A multinational replication of Prospect Theory

Preregistration – Data collection to begin July 2019

1 Kai Ruggeri, Columbia University & University of Cambridge

2 Sonia Alí, University of Sussex

3 Mari Louise Berge, Eötvös Loránd University

4 Giulia Bertoldo, University of Padova

5 Ludvig D Bjørndal, University of Oslo

6 Anna Cortijos-Bernabeu, University of Barcelona

7 Clair Davison, University of St Andrews

8 Emir Demić, University of Belgrade

9 Celia Esteban Serna, University College London

10 Maja Friedemann, University College London

11 Shannon P Gibson, Oxford Brookes University

12 Hannes Jarke, RAND Europe

13 Ralitsa Karakasheva, University of Nottingham

14 Peggah R Khorrami, Columbia University

15 Jakob Kveder, University of Ljubljana

16 Thomas Lind Andersen, Nordfyns Municipality

17 Ingvild S Lofthus, University of Oslo

18 Lucy McGill, Trinity College Dublin

19 Sahana K Quail, King's College London

20 Charlotte Rutherford, University of Cambridge

21 Felice L Tavera, University of Cologne

22 Nastja Tomat, University of Ljubljana

23 Chiara Van Reyn, KU Leuven

24 Bojana Većkalov, University of Belgrade

25 Keying Wang, University College London

26 Francesca Papa, Organisation for Economic Cooperation and Development

27 Enrico Rubaltelli, University of Padova

28 Sander van der Linden, University of Cambridge

29 Tomas Folke, University of Cambridge

Abstract

One of the most influential papers across all of the behavioural sciences is Kahneman and Tversky's 1979 article proposing Prospect Theory. This manuscript presented a series of binary financial choices with risk, ultimately concluding that behaviours deviate significantly from those presumed by prevailing theory at the time. In the forty years since publication, this study has had a remarkable impact on science, policy, and other real-world applications. At the same time, a number of critiques have been raised about its conclusions and subsequent constructs that were founded on it. In an era where such presumed canonical theories have drawn scrutiny for inability to replicate, we propose a multi-country study rerunning the same methods as in the original paper. This sample will test the original method, with modifications based on currency inflation and using all items for all participants in a random order. With a sample of over 2,000 individuals from at least 15 countries and languages using contemporary analytical approaches, we anticipate a general replication of findings from the 1979 study. This registered report presents a detailed justification for running this replication, an extensive protocol for collection and analysis, and anticipated findings in advance of work being undertaken.

Introduction

One of the most influential papers across all of the behavioural sciences is the 1979 *Econometrica* article by Daniel Kahneman and Amos Tversky, entitled *Prospect Theory: An analysis of decision under risk* (Kahneman & Tversky, 1979). The study was conducted with Israeli, Swedish, and American students and university faculty participants (item sample sizes between 64 and 141). The items followed a typical structure in decision-making research: binary financial choices with probabilistic outcomes. Across 20 items, in which various choices ('prospects') were presented in terms of value and probability, the authors established that patterns diverted substantially from the predictions of expected utility theory (EUT), the dominant descriptive theory at the time.

Prior to widespread acceptance of Prospect Theory, decision-making approaches largely accepted prevailing traditions on choice related to expected values, utility, and the axioms of rational behaviour. As such, choice was generally viewed from the assumption that rational individuals seek to optimize using stable algorithms tied to value, probability, and cumulative wealth. Based on the findings in their study, Kahneman and Tversky conclude that certainty in gains carry greater weight, whereas risks are preferable when losses are likely, and that small probabilities are often over-weighted. They argue that utility is more likely dependent on the change in assets relative to a reference point, and not on the absolute outcome asset, as previously believed. These decisions also appear to be influenced by how information is framed, meaning that the same values presented in different ways can elicit different choices from the same individuals, and therefore different patterns of behaviours across populations.

Critically, they convincingly argued that people were largely loss averse: that most people find losing a certain amount more aversive than gaining the same amount was appetitive. Thus, if given the choice between maintaining current wealth and accepting a 50-50 gamble of gaining $1000 or losing $1000, very few people take the gamble. Furthermore, the larger the sums involved, the fewer people would

accept it. Yet, when given the choice between a certain loss, and a gamble that could lead to no loss or a greater loss, most opt to take the gamble, seemingly to avoid the loss altogether.

It is difficult to overstate the level of influence that Prospect Theory has had on science, policy, management, financial services, government, and beyond. Shortly after Daniel Kahneman received the 2002 Nobel Prize for Economic Sciences, it was identified as the most influential theoretical framework in all of the social sciences (Mercer, 2005). It has since become the most cited economic paper and among the most-cited in psychological science (Simonsohn, 2014). Even the story around the relationship and collaborative partnership of Kahneman and Tversky became a bestselling book by award-winning author Michael Lewis.

Impact

The loss aversion component of Prospect Theory has provided an explanatory framework for understanding a broad range of behaviours (Barberis, 2013) in finance (Altman, 2010; Odean, 1998) and investment (Genesove & Mayer, 2001; Benartzi & Thaler, 1995), insurance (Johnson et al., 2013; Sydnor, 2010), and political conflict (Levy, 1996). Extensions of the work provided foundational insights to explain behaviour under riskless circumstances. Those include endowment effect (Thaler, 1980; Tversky & Kahneman, 1981; Köszegi & Rabin, 2006), status quo biases (Kahneman et al., 1991), and the hedonic impact ratings assigned to losses compared to gains (McGraw et al., 2010).

It has now been forty years since the initial manuscript was published. The findings of the original study became increasingly discussed in recent years in efforts to understand advances in behavioural sciences as well as in major world events. Perhaps the largest leap into such mainstream recognition when used as the core science behind Thaler and Sunstein's landmark writing on nudge theory, which popularized behavioural economics and revolutionised approaches to policy in organisations and governments

around the world (Thaler & Sunstein, 2009). Given this prominence, it is important to reassess the initial study to determine what outcomes appear if studied again today.

Debates and basis for replication

As should happen in science, with notoriety and acceptance has come a number of critiques of Prospect Theory, including from the original authors themselves. For example, Kahneman and Tversky realized that Prospect Theory struggles with prospects with large number of potential outcomes and only deals with risky choices (where probabilities are known) as opposed to uncertain choices (where they are not). Consequently, they updated their theory to deal with these shortcomings (Tversky & Kahneman, 1991).

A recent criticism levied against one of its primary tenets – loss aversion - claims that more rigorous evaluation of its findings and those that are borne from it, are the loss of loss aversion. In a narrative, non-systematic review, Gal & Rucker claim that current evidence does not support the notion that losses are more impactful than gains and the theory merely remains in strong support through persistent beliefs among scientists and that most conclusions present "inferential concerns" that can be reimagined (Gal & Rucker, 2018). For example, that people sell too quickly stocks that had early gains and hang on to those that lose value slowly not because of the disposition effect (based on loss aversion) as argued by Odean (1998), but simply because of perceived mean reversion on the part of the trader.

While those arguments were largely refuted in scientific and mainstream literature (Ritholtz, 2018), relevant debates on loss aversion raised in behavioural science literature continue. For example, Plott and Zeiler (2007) argue that changes in experimental conditions can significantly affect the magnitude of exchange asymmetries and willingness-to-accept/willingness-to-pay gaps, leading them to question the loss aversion interpretation of these effects. They suggest that the exchange asymmetries may be due to

(incorrectly) perceiving the object they were initially given as more valuable. (It should be noted that Kőszegi and Rabin (2006) counter that those findings are in fact consistent with loss aversion when the reference point is the decision-maker's expectations). Other examples of debates on loss aversion include:

- It is argued that loss aversion does not emerge in cases where gains and losses are repetitive or when amounts in question are not very high (Erev, Ert, & Yechiam, 2008) (McDermott, 2004).
- The reluctance to accept a lottery with equal chance of losing and winning the same amount of money may be a status quo bias rather than loss aversion (Gal, 2006).
- Loss aversion may be countered with perspective-taking, which suggests it is not a strong form of bias (Sokol-Hessner et al., 2009).
- If gains and losses are placed in comparable ranges, thereby modifying the subjective value of the loss, aversion to loss can reverse (Walasek & Stewart, 2015).
- Loss aversion may hold true in contained scenarios, but seems to dissipate in contexts where there may be a steady exchange of goods (Barberis, 2013).

Given the time that has passed since the original study, its position among the foundations of behavioural science canon, its widespread influence in science and policy (McDermott, 2004), and relevant critiques of its methods and conclusions, Prospect Theory deserves an unbiased reassessment. This is particularly critical in light of concerns about replicability in behavioural science as well as the translatability of findings between locations (Klein et al., 2014). It has been convincingly argued that direct replication by other laboratories is the best way to provide evidence for the robustness of psychological effects (Simons, 2014). While loss aversion itself is rarely refuted on the whole, the idea that this behaviour only appears in very specific situations could be investigated by a study using non-specific samples across cultures.

Since multiple theories that were presumed canon have failed to replicate at a convincing level (Klein et al., 2018), it is important to test those well-known beliefs even if (and perhaps especially if) they are generally accepted by the scientific community. In other words, it is particularly important to attempt replication in areas where effects are presumed robust, which certainly applies to loss aversion (Walasek & Stewart, 2015). This may be specifically true in a "hot" scientific field (such as behavioural economics, which was largely spurred by Prospect Theory), a concern raised as Corollary 6 by Ioannidis (2005): more popular topics are more likely to produce false results.

This study aims to bring fundamental observations leading to the formation of Prospect Theory in line with the standards of replicability in behavioural science in 2019. Our approach emphasizes what is critical for behavioural science: replication of a major study. A failed replication would have seismic implications for behavioural science, but a (in our view, likely) successful replication would still hold scientific value. Not only would it strengthen our confidence in core assumptions about decision-making, given the multi-national nature of the replication we will be able to document variability between locations and languages (or the lack thereof) in key aspects of financial decision-making under risk.

Methods

This trial will involve - as closely as possible - the direct replication of items used in the original paper on Prospect Theory. Original conclusions and contemporary analytical approaches will be tested through a combination of descriptive and inferential analyses. In most cases, these will treat findings from 1979 as the baseline. Further steps will assess deviation from that model across groups and locations. Additional items are also introduced to account for various demographic factors as well as knowledge of the hypothesized effects. In-principle ethical approval has been provided by the Centre for Business Research in the Judge Business School at the University of Cambridge, where this project will be

primarily housed. A final approval will be provided following review, after the registered protocol is final.

Instrument

In order to closely replicate the same procedure, measures will largely repeat the original approach from the 1979 publication. We will use the same items as published in Kahneman and Tversky (1979), but will exclude the two verbal travel items and the verbal probabilistic insurance item (5, 6, & 9 in the original publication). The travel items, which entailed a choice between having a chance to travel to England, France, and Italy, or having a certain trip to England, will be excluded. Their subjective utility might differ markedly between countries, which adds a needless interpretive burden to this multi-country replication. Additionally, given the events of recent years it is unclear whether a "certain trip to England" would have a positive utility for all participants. The probabilistic insurance item was excluded because the wording of the item was long and convoluted, creating concern that it more likely tests reading comprehension rather than theoretically relevant constructs.

The full set of items in English are appended to this manuscript. Forward and back translation will be used for all measures, with adjustments to local currency, but this is a relatively simple process given the nature of the instrument.

Procedure

All participants within each testing location will complete the exact same survey of 27 total items, including demographics. The only variation will be to account for language and currency differences,

plus order (see *Accounting for order effects*). After providing informed consent, participants will respond to 17 choices with risk from the original study. The order of the choice items and response options will be completely randomised. This is similar in spirit to the 1979 paper, though that randomisation in the original was more limited, as the printed nature of surveys only allowed them to include a few different presentation orders.

The original paper presented each participant multiple items in a pseudo-randomised order, but no participants encountered all of the items in the paper and it is not always clear from the manuscript which items belonged to the same survey. In this replication, all participants will encounter all items, which should not be a confounder as participants in the original study responded to multiple items, so the experience of completing the survey will be similar. Our primary aim is to test if original conclusions hold with only minor improvements to the original method.

Nine demographic measures will come after decision-making items to avoid any results being influenced by stereotype threat. Participants must complete the survey within three median absolute deviations of the median completion time or their data will be excluded (Leys, 2013). This criterion is to ensure that they are genuinely answering the survey, not clicking rapidly to complete or multi-tasking such that answers are not given in reasonable succession. A "paying attention" item will also be randomly included – anyone that answers this incorrectly will be excluded. This item gives the simple instruction of "Do not choose either option, just proceed to the next question." Two options are presented, either of which, if answered, will immediately exclude the participant by ending the survey. The options are between a guaranteed gain of 10,000 or a 99% chance of losing 5,000, which means participants that are truly reading the options should immediately notice an obvious departure from other items.

Additional items for risk were considered to counter-valence the original methods, but these were decided against as requiring an entirely new theoretical basis.

At the end of the survey, we will ask participants if they are familiar with the concept of loss aversion, as a proxy for general awareness of behavioural economics. We have no hypothesis regarding how awareness of loss aversion might bias participants to answer in line with expectations (i.e., similar patterns as the 1979 results) or counter to them. However, we will test if any differences exist between those who claim they are familiar with the concept with those that are not.

Financial values

To replicate as closely as possible the weight of values in the original study in a contemporary setting, two approaches will be applied to account for inflation, currency exchange, and location. The primary approach will be adjusting items directly toward median household incomes in each location for June 2019 relative to the original study, when the median net household income for Israel was about 3,000 Israeli pounds per month. For example, an item that was 2000 Israeli pounds in 1979 was ⅔ of the 3,000 Israeli pounds per month median income. For 2019 in the United States, median income is about $5,400 per month, so the same item would use $3,600 for US participants in the replication. However, as this produces values that may make participants suspicious about prospects, particularly where median values do not convert into whole numbers, slight adjustments to get rounded values will be used. All within-item prospects will retain the same relative values as in the 1979 instrument.

For some countries, it may be necessary to apply inflation rates to the original values, then calculate current exchanges to local currencies. Ideally, this would have been done in the reverse order, but multiple currencies used in this study did not exist in 1979 (e.g., euros, Serbian dinars). This approach

will only be applied to a small number of countries to account for potential effects of rising income inequality (i.e., uneven relative value of median incomes). An example table of conversions is included in the appendix. This approach has ultimately been decided on the basis it is more important to expose participants to choices representing the same wealth as the original study, rather than using the exact same numbers.

It is worth noting that all items involve hypothetical monies only, in line with the original study. The lead author and several collaborators have recently completed a multi-country study showing answers do not change substantially between hypothetical items and those involving real money. This comparability has also been shown in other work (Kühberger et al., 2002; Beattie & Loomes, 1997; Wiseman & Levin, 1996). Therefore, there was no need to introduce items with real-world gains and losses for validation.

Demographics

The primary demographics of interest will be nationality, year of birth, sex, income, educational attainment, four measures of current financial circumstances and behaviours (strain, recent changes, investments, debts). We do not anticipate substantial differences between any groups, though we anticipate some variability is likely. Language of assessment will be included in the data.

Sample

There will be two tracks for registering participants in the study. The first will be direct contact with convenience samples for general testing of the procedure, followed by a sample recruited through an online platform. This generally follows the replication approach used in the ManyLabs trials (Klein et al., 2014; Klein et al., 2018), noting that we will intentionally not utilise psychology student participant pools (Ebersole et al., 2016).

Prior to online collection, all project members will conduct convenience sampling using the same instrument for a minimum of 73 participants. This number has been chosen because it is above half of the sample that will be collected online to meet necessary power levels (see Power Calculation), larger than the smallest sample (64) in the 1979 study, and also in line with the smallest laboratory sample in the major 2014 multinational replication trial. Convenience sampling in this case is intentional to assess potential differences between an online sample and one with engrained, systemic bias. Analyses will include composite and separated samples, but with the wider view that if loss aversion is indeed present, and the tenets of Prospect Theory intact, original 1979 results should be generally observed across populations (even with some variability).

The online sample will be paid for their participation, and largely recruited via Prolific Academic, where available. Other paid recruitment platforms will also be used.

At present, participants will be recruited in the languages (countries) listed below. There is no systematic method for inclusion beyond the collaborators that have volunteered to participate. While there is a noted skew towards Euro-American regions, the generally random nature of inclusion is helpful for avoiding some level of systematic bias for participants.

- English (United Kingdom, United States, Ireland, Hong Kong, Australia, Singapore)
- Italian (Italy)
- Spanish (Spain)
- German (Germany, Austria, Switzerland)
- Serbian (Serbia)

- Slovenian (Slovenia)
- Hungarian (Hungary)
- Norwegian (Norway)
- Swedish (Sweden)
- Chinese (Hong Kong, China)
- Bulgarian (Bulgaria)
- Dutch/Flemish (Belgium, Netherlands)
- Danish (Denmark)

Additional countries where Spanish (Colombia, Mexico, United States), Italian (Switzerland), German (Switzerland), French (Belgium, France, Switzerland), Russian (Russia, Kazakhstan), and Arabic (Lebanon, Qatar), are commonly spoken may also be further included. However, the collaborators responsible for these countries could not commit participation until after a final protocol is confirmed. Data from these additional locations will be presented under exploratory analyses, though analytical procedures will reflect the core protocol presented here.

All data collection will emanate from a single institutional account, with data collected exclusively online; no *in situ* testing will take place.

Power and error

We will first test if effects are similar to the 1979 trial across items, looking for significant deviation between replication findings and original findings. We will assess this for all countries and groups, as well as in composite. Such descriptive approaches will yield a large number of outputs, and we expect

any substantial differences between the original trial and replication to be spurious (i.e., across all group comparisons, 5% or fewer should yield a significant difference from the original trial).

Though the original paper only reports chi-square tests that compare the response distribution of each item relative to a balanced null distribution, Kahneman and Tversky's theoretical argument primarily relies on contrasts in response distributions between different items (see Appendix 2). The smallest of these reported contrasts are between items 4 and 8, where 65% and 42% chose option A, respectively. This gives an odds ratio of 2.56. To have 95% power to detect such an effect at an alpha value of .05 requires at least 120 participants per cell. Thus, in order to account for dropouts and exclusions, we will aim to test at least 145 participants per country, which is larger than the largest sample size reported in the original study (141). Because all participants will complete all items, this means that the number of participants per cell will be 145 assuming no-drop outs, and will almost certainly be greater than the threshold of 120 per cell required by the power analysis. Additionally, because the key analyses for hypothesis testing will be hierarchical regression models (see the fourth set of analyses specified in the analysis plan), true power will be even higher than implied by the power analysis, where each country is treated as independent. However the power boost caused by partially pooling the data depends on the heterogeneity of effects between countries (Gelman & Hill, 2012), which is hard to estimate at this stage.

This also matches the approach presented by ManyLabs (Klein, 2014) in replicating multiple psychological studies (one of which included gain-loss framing items published by Kahneman and Tversky in 1981). However, where ManyLabs found partial method replication attempts were not impacted by setting (Owens, 2018), we will test a single method comprehensively between settings.

Our approach ensures that the sample size for each country will give sufficient power for testing within locations as well as in composite. Note that because we aim to collect data from a minimum of 15 countries, the total sample will be at least 2,910, which would give us a power of 1 to detect an odds ratio at 2.56 (smallest 1979 effect) at an alpha level of .0001. All sample size calculations and power calculations are based on the bpower function in the Hmisc package in R.

We note that our emphasis will primarily consider Type 1 and Type S errors. We will report, but not focus heavily, on potential Type M errors (i.e., the factor by which a statistically significant effect size in the replication overestimates the plausible effect size) (Gelman & Carlin, 2014).

Analysis

To fully explore the replicability of the effects reported in the original prospect theory we will run four sets of analyses, of increasing complexity. All analysis code and data will be made available on GitHub.

First set: Copying the original analyses and running descriptives for the replication data

In the first set of analysis, we will compare the results to the first participants completing the study, so that our sample sizes equal the original study sample (for example, if there were 72 participants responding to one item in 1979, we will analyse the first 72 participants who responded to that item in 2019). Here we will exactly replicate the analysis style of Kahneman and Tversky's paper in that we will run chi-square tests comparing each items response distribution with a null-distribution where both options were preferred equally. Following this, we will look at the effect-sizes for each item, expressed as proportion of choices for option A, split by country and recruitment method, and plot them relative to the original proportions, similar to figure 1 in Klein et al (2014). These first set of analyses are largely

descriptive and directly compare our results to 1979, in a similar format to which they were presented there.

Second Set: Quantifying variability in the effect sizes by bootstrap

Next, we will explore the range of plausible effect-sizes for each item by country and collection method by bootstrap analyses (Kuonen, 2018). Here we will record the distribution of effects for 100,000 random samples per item, and report what proportion of samples fall below the significance threshold in the predicted direction. The seed for the random sampling will be set to 357, to ensure that others can faithfully recreate our result. This seed was determined by running the numpy random choice function in Python, to randomly select a number between 0 and 1000.

Third Set: Quantifying the effects of demographic variables on choice through Bayesian Hierarchical modelling

Third, just as we obtained a plausible range of effect sizes from bootstrapping, we can use Bayesian parameter estimation to get a sense of plausible effects. The relative benefit of the Bayesian approach is that it does not only capture the uncertainty in our estimates, but also allows us to estimate how much variation in the effect can be attributed to differences in nationality and other demographic factors, as well as interactions between demographic factors, whilst accounting for the hierarchical structure of the data (Gelman & Hill, 2012). We will use Bayesian logistic regressions to derive both global and national effect-size estimates for each item, as a function of the demographic variables we have access to. Hyperpriors will be set to Normal(0, 1) for means and Exponential(1) for standard deviations in line with McElreath's recommendations for logistic hierarchical models (McElreath, 2018). All models will include age, gender, awareness of loss aversion, and economic circumstance as predictors of choice. Both intercepts and demographic effects will be partially pooled, meaning that they will be allowed to differ

by country, but that they will be pulled towards a global trend by a shrinkage parameter. Main results will cover the posterior probability of the fixed effect, and the posterior probability of between country variances, for each demographic variable. Plots that show the posterior probability of each demographic effect per question per country will be available in the supplement.

Formally these models will have the following form:

$A_i \sim \text{Binomial}(n_i, p_i)$

$$\text{logit}(p_i) = \alpha_{country[i]} + Sex_i\beta 1_{country[i]} + Age_i\beta 2_{country[i]} + Income_i\beta 3_{country[i]} + Education_i\beta 4_{country[i]} + FinCirca_i\beta 5_{country[i]} + FinCircb_i\beta 6_{country[i]} + LossAwareness\beta 7_{country[i]} + LossIntuition\beta 8_{country[i]}$$

$\alpha_{country} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha)$

$\mu_\alpha \sim \text{Normal}(0, 1)$

$\sigma_\alpha \sim \text{Exponential}(1)$

For all β:

$\beta_{country} \sim \text{Normal}(\mu_\beta, \sigma_\beta)$

$\mu_\beta \sim \text{Normal}(0, 1)$

$\sigma_\beta \sim \text{Exponential}(1)$

Where A signifies choosing option A, Sex is dummy coded as 0 for women and 1 for men, LossAwareness is dummy coded as 1 if participants report awareness of loss aversion and 0 otherwise, LossIntuition is

dummy coded as 0 for those who correctly intuit loss aversion (but report no knowledge of the effect) and 0 otherwise, Age and Income is z-scored within each country, and Education and the FinancialCircumstance variables are ordered factors.

Fourth Set: Testing the core predictions of Kahneman and Tversky in a Bayesian Framework

Though Kahneman and Tversky never compared items statistically, the bulk of their argument rests on comparisons between choice patterns for two items (See appendix 2 for a list of these comparisons, and which effect they evaluate). As such, these comparisons are much more theoretically relevant than the comparison between each item and a null-model, as is presented in the original paper. To quantify these effects we will run hierarchical Bayesian models, similar to those described above, with the distinction that we will predict choices for two items at once with a dummy variable coding for item. This dummy variable will be nested in country. This will allow us to directly quantify a global estimate of the effect sizes for the behavioural effects evaluated in the 1979 paper (the certainty effect, the reflection effect, the isolation effect, the overweighting of small probabilities, and range adaptation) as well as estimate the extent to which they vary between countries.

Formally these models will have the following form:

$$A_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \alpha_{country[i]} + \text{Item}_i\beta 1_{country[i]} + \text{Sex}_i\beta 2_{country[i]} + \text{Age}_i\beta 3_{country[i]} + \text{Income}_i\beta 4_{country[i]} + \text{Education}_i\beta 5_{country[i]} + \text{FinCirca}_i\beta 6_{country[i]} + \text{FinCircb}_i\beta 7_{country[i]} + \text{LossAwareness}\beta 8_{country[i]} + \text{LossIntuition}\beta 9_{country[i]} + \text{ItemDistance}\beta 10_{country[i]} + \text{ItemDistance:Item}_i\beta 11_{country[i]} + \text{CountryReferenceValue}\beta 12 + \text{CountryReferenceValue:Item}_i\beta 13$$

$\alpha_{country} \sim Normal(\mu_\alpha, \sigma_\alpha)$

$\mu_\alpha \sim Normal(0, 1)$

$\sigma_\alpha \sim Exponential(1)$

For all β:

$\beta_{country} \sim Normal(\mu_\beta, \sigma_\beta)$

$\mu_\beta \sim Normal(0, 1)$

$\sigma_\beta \sim Exponential(1)$

Where A signifies choosing option A, Item is a dummy coded as 0 for the first item and 1 as its contrast. This variable captures the core behavioural effects prospect theory attempts to account for. ItemDistance codes for the distance between the two contrast items, i.e. 0 if the items are adjecent, 1 if there is one item between them and so forth. These items are included to test for proximity effects, e.g. it is conceivable that the framing effect is moderated if the same item in the gain frame and loss frame are presented immediately after each other relative to some distance apart. Apart from testing for these effects in formal modelling we will plot these effects as well, to visually examine the functional form of these proximity effects. The CountryReferenceValue is the reference value (median income in local currency) used for a specific country, ranging between 1,200 and 60,000. This is included as a predictor to test if the magnitude of the numbers somehow influence how people make choices about risk, for example by causing artificial invariances with people ignoring values below a specific decimal points. We do not expect to find any dramatic effects here, but in case we do it is worth noting that two countries

in this study is using the same reference value as the original PT study (3000), with one of these countries having a mean monthly income slightly above 3000 in the local currency and the other one slightly below. The remaining variables coded the same as the previous model.

We will aim to present a number of visualisations that may better highlight general patterns than individual analyses presented in numeric prose. These will include producing a parallel for the "s-curve" as depicted in the 1979 manuscript, using the available data to show what curves appear as with selection proportions included. These will be produced in composite as well as, if possible, separated or with overlays to show if curves differ by group or location. Other visuals, as described, will be similar to those presented in the Klein et al. replication trial (2014), which appear similar to meta-analyses by country, testing type, and other breakdowns mentioned.

We will also create an interactive online tool that lets people explore the differences between countries, in the simplest case an open access spreadsheet, in the most complex case an interactive dashboard. We will aim for this tool to allow, at a minimum, to analyse random subsets of the current data equivalent to the 1979 sample size as well in composite by countries, overall, and by groups.

Limitations

First, it is important to highlight that this study is testing whether the original 1979 work by Kahneman and Tversky replicates in a modern sample from multiple countries. This is in itself a critical goal, but will not address all criticisms linked to Prospect Theory or loss aversion. However, replication is a critical first step in this process, in the sense that though a successful replication would not neutralise all possible criticisms, yet a failed replication would suggest serious issues with the original theory.

To avoid future debate on this topic: we are fully aware that arguments challenging Prospect Theory and loss aversion did not necessarily or categorically challenge the original method, but rather focus on context, interpretations, and conclusions (e.g., Katsikopoulos & Gigerenzer, 2008). We highlight that the purpose here is to ensure that original findings do in fact replicate widely, and would offer further insights about its conclusions only after that were confirmed or refuted. Concurrently, we will also be able to test potential order effects of choice as opposed to one-off behaviours, though without accumulation as items progress, we cannot assess sequential choice patterns comprehensively. It is also unlikely we will be able to test if one theoretical framework (e.g., status quo bias) supersedes or dominates Prospect Theory using the data we generate.

Using the two sampling tracks – convenience and pay-sourced - does also leave it open to concerns about participation. We will be transparent about those limitations once the final data are available and full sampling is known.

One pattern that has largely been ignored in all work in decision-making with risk is the minority group outcome, and loss aversion is perhaps the strongest instance of this. Where a clear majority may consistently choose the certain, lower value, there are always *some* that choose the risky option or accept a certain loss of lower value. We are not able to spend more time exploring these in this study, but will analyse those patterns in the future while also making the data available. We simply highlight this here to suggest that there may be the archetype *exception that proves the rule*, and should be analysed in their own right. One criticism of loss aversion conclusions implies that because exceptions exist, the theory must be incorrect. This is fallacy, as has been pointed out in various literature (Ritholtz, 2018): sometimes taking the risk is something we prefer to do, such as spending a weekend in Las Vegas, playing the lottery with colleagues, or waiting for the next train hoping it might be the express. We will not have data to cover all such arguments.

Discussion (Planned outline only)

With an attempt to replicate, it is possible to test both the original evidence used to form the theory, and guide continued debate about conclusions derived therein. Failure to replicate, in this way, avoids irrelevant conclusions and hopefully provides meaningful direction for next steps. In advance of data collection, we provide a list of our primary topics for discussion to address these points in the final manuscript. This will include:

1. Critical insights related to methods, procedure, and sample, particularly if any steps deviate from plan, or if any relevant issues arise.
2. General summary of all items and patterns of choice and risk.
3. Regional variability, particularly focusing on any significant deviations from 1979 results by country.
    a. If particularly relevant, we will present any variability within or between demographic categories, but this will largely be included in the supplement or published elsewhere.
4. A chronological (using 1979 order) review of each item
5. Do any established critiques, particularly those from the original authors, seem to hold?
6. General statement on whether Prospect Theory (and its successors) and loss aversion appear to replicate.
7. Potential of error and noise in findings, both within the sample and in comparison to 1979 results.

Timeline

We expect to begin data collection by mid-July 2019, which should last approximately 15 days. We will have pre-scripted analyses in R software for immediate analysis, which should last another 15 days. We anticipate submission of the full manuscript by 27 August, 2019.

Conclusion

We come into this study unbiased and without vested interest in the results of the trial. While we acknowledge prior commitment to this field of work and use of the theory under question, the recent critique presented sufficient impetus to directly test long-held notions. Such challenges are critical for ensuring scientific quality, no matter how widely accepted conclusion may be.

In spite of this, we strongly expect a full and comprehensive replication of findings, if not universal. Furthermore, in the absence of a strong counter-argument from critiques mentioned, we also anticipate loss aversion, framing effects, and overweighting small probabilities will only remain largely intact and among the best-established empirical conclusions in the social and behavioural sciences. We present threshold for making these determinations and the conclusion of this study, yet anticipate the notion that loss aversion being fallacy to be broadly rejected. In either event, we will present outcomes as they appear in these planned analyses given their considerable value to individual decision-making through to population well-being.

References

Altman, M. (2010). Prospect theory and behavioral finance. In H. K. Baker & J. R. Nofsinger (Eds.), *Behavioral Finance: Investors, corporations, and markets* (Vol. 6)., 191-209. Wiley.

Barberis, N. C. (2013). Thirty years of Prospect Theory in economics: A review and assessment. *Journal of Economic Perspectives*, *27*(1), 173-96.

Beattie, J., & Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty, 14(2)*, 155-168.

Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, *110*(1), 73-92.

Ebersole, Charles R., Olivia E. Atherton, Aimee L. Belanger, Hayley M. Skulborstad, Jill M. Allen, Jonathan B. Banks, Erica Baranski et al. (2016) "Many Labs 3: Evaluating participant pool quality across the academic semester via replication." *Journal of Experimental Social Psychology,* 67, 68-82.

Erev, I., Ert, E., and Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making, 21*, 575-597.

Gelman, A., & Hill, J. (2012). *Data analysis using regression and multilevelhierarchical models* (Vol. 1). New York, NY, USA: Cambridge University Press.

Gal, D. (2006). A psychological law of inertia and the illusion of loss aversion. *Judgment and Decision Making, 1*, 23–32.

Gal, D., & Rucker, D. D. (2018). The loss of loss aversion: Will it loom larger than its gain? *Journal of Consumer Psychology*, *28*(3), 497-516.

Gelman, A., & Carlin, J., (2014). Beyond power calculations: Assessing Type S (Sign) and Type M (Magnitude) errors. *Perspectives on Psychological Science, 9(6),* 641-651.

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics, 116(4)*, 1233-1260.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Johnson, E. J., Hassin, R., Baker, T., Bajger, A. T., & Treuer, G. (2013). Can consumers make affordable care affordable? The value of choice architecture. *PloS One*, *8*(12), e81521. doi:10.1371/journal.pone.0081521

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability. *Social Psychology, 45,* 142-152. https://doi.org/10.1027/1864-9335/a000178.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443-490.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133-1165.

Kühberger, A., Schulte-Mecklenbeck, M., & Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes, 89(2)*, 1162-1175.

Kuonen, D. (2018). An introduction to bootstrap methods and their application. *WBL in Angewandter Statistik ETHZ 2017*, *19*, 1-143.

Levy, J. S. (1996). Loss aversion, framing, and bargaining: The implications of prospect theory for international conflict. *International Political Science Review*, *17*(2), 179-195.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766.

McDermott, R. (2004). Prospect theory in political science: Gains and losses from the first decade. *Political Psychology, 25*(2), 289-312.

McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkade, D. (2010). Comparing gains and losses. *Psychological Science*, *21*(10), 1438-1445.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Mercer, J. (2005). Prospect theory and political science. *Annual Review of Political Science*, 8, 1-21.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance, 53,* 1775-1798.

Owens, B. Replication failures in psychology not due to differences in study populations. *Nature News* (19 November, 2018).

Plott, C., & Zeiler, K. (2007). Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory? *American Economic Review, 97*(4): 1449 – 66.

Ritholtz, B. (2018, August). A Challenge to the Biggest Idea in Behavioral Finance. *Bloomberg Opinion - Finance*.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76-80.

Simonsohn, U. (2014). '[15] Citing Prospect Theory'. *Data Colada, February 10. Available at http://datacolada.org/15*.

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*, *106*(13), 5035-5040.

Sydnor, J. (2010). (Over) insuring modest risks. *American Economic Journal: Applied Economics*, *2*(4), 177-99.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1(1),* 39-60.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

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

Katsikopoulos, K. V., & Gigerenzer, G. (2008). One-reason decision-making: Modeling violations of expected utility theory. *Journal of Risk and Uncertainty*, *37*(1), 35-56.

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.

Appendix 1. Consent form

Information and Consent

The following survey asks questions about your preferences regarding financial choices. You will be presented with a series of questions with various options where you have to pick between a risky and certain outcomes. There are no right or wrong answers to these questions, just pick whichever option you prefer. Afterwards, you will be asked some questions about your background. These are broad questions that do not allow you to be identified personally. The information we collect from you will only be used for the purposes of academic research. No one will contact you after.

The entire survey should take about three to four minutes. If you have any questions or concerns, please contact Dr Kai Ruggeri at dar56@cam.ac.uk.

By participating, you confirm the following:

- *you are at least 18 years of age*
- *you are a permanent resident of COUNTRY WHERE ADMINISTERED*
- *you consent to participate in the study*
- *you consent to your data used for the research described above*
- *you understand that your participation is completely anonymous*
- *you understand that you may stop at any time.*

If this is correct, please choose the option below and proceed to the survey.

○ I confirm these are accurate, that I understand the terms, and agree to participate in this study.

Appendix 1a. Original Prospect Theory items to be replicated

Note that the values in the original paper denoted Israeli pounds, Kahneman and Tversky state that the median monthly income for a family at the time of writing was about 3000 Israeli pounds. We use that as a reference so that the values hold the same relation to the median incomes in the countries we are testing (see Appendix 3).

1. Which option do you prefer?
    a. A 33% chance at 2500, a 66% chance at 2400, or a 1% chance of 0
    b. Guaranteed 2400
2. Which option do you prefer?
    a. 33% chance of 2500 (67% chance of 0)
    b. A 34% chance of 2400 (66% chance of 0)
3. Which option do you prefer?
    a. An 80% chance of 4000 (20% chance of 0)
    b. 100% guarantee of 3000
4. Which option do you prefer?
    a. An 20% chance of 4000 (25% chance of 0)
    b. 25% chance of 3000 (75% chance of 0)
5. Which option do you prefer?
    a. A 45% chance of 6000 (55% chance of 0)
    b. 90% chance of 3000 (10% chance of 0)
6. Which option do you prefer?
    a. A 0.1% chance of 6000 (99.9% chance of 0)
    b. 0.2% chance of 3000 (99.8% chance of 0)
7. Which option do you prefer?
    a. A 80% chance of losing 4000 (20% chance of losing 0)
    b. A 100% guarantee of losing 3000
8. Which option do you prefer?
    a. A 20% chance of losing 4000 (80% chance of losing 0)
    b. A 25% chance of losing 3000 (75% chance of losing 0)
9. Which option do you prefer?
    a. A 45% chance of losing 6000 (55% chance of losing 0)
    b. A 90% chance of losing 3000 (10% chance of losing 0)
10. Which option do you prefer?
    a. A 0.1% chance of losing 6000 (A 99.9% chance of losing 0)
    b. A 0.2% chance of losing 3000 (A 99.8% chance of losing 0)
11. Imagine you are playing a game with two levels, but you have to make a choice about the second level before you know the outcome of the first. At the first level, there is a 75% chance that the game will end without you winning anything, and a 25% chance that you will advance to the second level. What would you choose in the second level?
    a. An 80% chance of 4000 (20% chance of 0)
    b. A 100% guarantee of 3000
12. Imagine we gave you 1000 right now to play a game. Which option would you prefer?

a. A 50% chance to gain an additional 1000 (50% chance of gaining 0 beyond what you already have)
b. A 100% guarantee of gaining an additional 500

13. Imagine we gave you 2000 right now to play a game. Which option would you prefer?
    a. A 50% chance you will lose 1000 (50% chance of losing 0)
    b. A 100% chance you will lose 500
14. Which option do you prefer?
    a. A 25% chance of 6000 (75% chance of 0)
    b. A 25% chance of 4000 (75% chance of 2000)
15. Which option do you prefer?
    a. A 25% chance of losing 6000 (75% chance of losing nothing)
    b. A 25% chance of losing 4000 (75% chance of losing 2000)
16. Which option do you prefer?
    a. A 0.1% chance at 5000 (99.9% chance of 0)
    b. A 100% guarantee of 5
17. Which option do you prefer?
    a. A 0.1% chance of losing 5000 (99.9% chance of losing nothing)
    b. A 100% guarantee of losing 5

Attention item (Sixth item for all participants; Not randomised with other items)

18. Do not choose either option, just proceed to the next question.

    a. 100% chance of 10,000

    b. 99% chance of losing 5,000

Appendix 1b: New items and demographic indicators

19. What year were you born? [Dropdown list from 1901-2001]

20. Which do you identify as?

Male Female Other _____ Prefer not to answer

21. What was your total personal income (after taxes) in 2018? Please use your local currency.

22. In the past month, your financial situation…

(1) Became much worse (2) Became a little worse (3) Has not changed
(4) Became a little better (5) Became much better

23. How did you handle your bills for the last month?

(1) I could not pay all my bills last month
(2) I had to take on a lot of debt to pay bills last month
(3) I had to take on some debt to pay them
(4) I paid everything but saved nothing
(5) I paid everything and saved some
(6) I paid everything, saved some, and paid down other debts (liked credit cards, cars, mortgage)
(7) Everything is paid off and I have enough to save, invest, or spend freely

24. What is your highest level of education completed?

Primary school Secondary school (high school)
Vocational school College or university (Bachelors or equivalent)
Graduate (Masters or equivalent) MBA Doctoral degree

25. In a given month, what is your typical credit/debt situation? Do not include rent/mortgage or functional bills such as auto payments or utilities (e.g., electricity).

(1) My debts increase beyond what I earn in a given month
(2) I generally pay off only/mostly the interest but not the balance
(3) I make the minimum payment and the interest is charged
(4) I pay off much of the balance
(5) I pay off any balances every month
(6) I never purchase anything on credit or take on debt

26. Do you earn additional income from any of the following?

Private investments (stocks, mutual funds) Property that you rent out
Ownership in a business other than where you work A retirement account paid into regularly
I do not have any of these

27. True or false: Research shows that people dislike losing more than we like winning.

True False
I do not know, but my **guess** is that it is **true**
I do not know, but my **guess** is that it is **false**

Appendix 2: List of results and key contrasts from prospect theory paper

Item	Prospect A						Prospect B						EV A	EV B	Contrast Item(s)	Effects evaluated	n	Proportion Choosing Prospect
	Value	Probability	Value	Probability	Value	Probability	Value	Probability	Value	Probability	Value	Probability						
1	2500	0.33	2400	0.66	0	0.01	2400	1					2409	2400	2	Certainty Effect	72	18
2	2500	0.33	0	0.67			2400	0.34	0	0.67			825	816	1	Certainty Effect	72	83
3	4000	0.8	0	0.2			3000	1					3200	3000	4, 7	Certainty Effect, Reflection Effect	95	20
4	4000	0.2	0	0.8			3000	0.25	0	0.75			800	750	3, 8, 11	Certainty Effect, Reflection Effect, Isolation Effect	95	65
5	6,000	0.45	0	0.55			3000	0.9	0	0.1			2700	2700	6, 9	Overweighting of small probabilities, Reflection Effect	66	14
6	6000	0.001	0	0.999			3000	0.002	0	0.998			6	6	5, 10	Overweighting of small probabilities, Reflection Effect	66	73
7	-4000	0.8	0	0.2			-3000	1					-3200	-3000	8, 3	Certainty Effect Losses, Reflection Effect	95	92
8	-4000	0.2	0	0.8			-3000	0.25	0	0.75			-800	-750	7, 4	Certainty Effect Losses, Reflection Effect	95	42
9	-6000	0.45	0	0.55			-3000	0.9	0	0.1			-2700	-2700	10, 5	Overweighting of small probabilities Losses, Reflection Effe	66	92
10	-6000	0.001	0	0.999			-3000	0.002	0	0.998			-6	-6	9, 6	Overweighting of small probabilities Losses, Reflection Effe	66	30
11	4000	0.2	0	0.8			3000	0.25	0	0.75			800	750	4	Isolation Effect	##	22
12	2000	0.5	1000	0.5			1500	1					1500	1500	13	Framing Effect	70	16
13	2000	0.5	1000	0.5			1500	1					1500	2500	12	Framing Effect	68	69
14	6000	0.25	0	0.75			4000	0.25	2000	0.25	0	0.5	1500	1500	15	Range adaptation	68	18
15	-6000	0.25	0	0.75			-4000	0.25	2000	0.25	0	0.5	-1500	-1500	14	Range adaptation	64	70
16	5000	0.001	0	0.999			5	1					5	5	17	Overweighting of small probabilities, Reflection Effect	72	72
17	-5000	0.001	0	0.999			-5	1					-5	-5	16	Overweighting of small probabilities Losses, Reflection Effe	72	17

Appendix 3: Example table of values to be used by country

1979 Values (Israeli pounds)	% of 1979 reference (3,000)	United States	Germany	United Kingdom	Norway	Ireland	Belgium	Australia	Spain	Italy	Hungary
2500	83.3%	5000	3000	2000	35000	2500	1500	3000	2000	1750	250000
2400	80.0%	4800	2880	1920	33600	2400	1440	2880	1920	1680	240000
4000	133.3%	8000	4800	3200	56000	4000	2400	4800	3200	2800	400000
3000	100.0%	6000	3600	2400	42000	3000	1800	3600	2400	2100	300000
6000	200.0%	12000	7200	4800	84000	6000	3600	7200	4800	4200	600000
1000	33.3%	2000	1200	800	14000	1000	600	1200	800	700	100000
500	16.7%	1000	600	400	7000	500	300	600	400	350	50000
5	0.167%	10	6	4	70	5	3	6	4	3.50	500
1979 Values (Israeli pounds)	% of 1979 reference (3,000)	Serbia	Switzerland	Austria	Hong Kong	China	Singapore	Slovenia	Sweden	Denmark	Bulgaria
2500	83.3%	50000	5000	2500	22500	6250	7500	1500	25000	25000	1000
2400	80.0%	48000	4800	2400	21600	6000	7200	1440	24000	24000	960
4000	133.3%	80000	8000	4000	36000	10000	12000	2400	40000	40000	1600
3000	100.0%	60000	6000	3000	27000	7500	9000	1800	30000	30000	1200
6000	200.0%	120000	12000	6000	54000	15000	18000	3600	60000	60000	2400
1000	33.3%	20000	2000	1000	9000	2500	3000	600	10000	10000	400
500	16.7%	10000	1000	500	4500	1250	1500	300	5000	5000	200
5	0.167%	100	10	5	45	12.50	15	3	50	50	2

The full table for all countries will be included in the final manuscript.