WELL EXCUSE ME! REPLICATING AND CONNECTING EXCUSE-SEEKING BEHAVIORS

BEATRIZ AHUMADA, YUFEI CHEN, NEERAJA GUPTA, KELLY HYDE, MARISSA LEPPER, WILL MATHEWS, NEIL SILVEUS, LISE VESTERLUND, TAYLOR WEIDMAN, ALISTAIR WILSON, K. PUN WINICHAKUL, AND LIYANG ZHOU

Abstract. Excuse-seeking behavior that facilitates replacing altruistic choices with self-interested ones has been documented in several domains. In a laboratory study, we replicate three leading papers on this topic: Dana et al. (2007), and the use of information avoidance; Exley (2015), and the use of differential risk preferences; and Di Tella et al. (2015), and the use of motivated beliefs. The replications were conducted as part of a graduate course, attempting to embed one answer to the growing call for experimental replications within the pedagogic process. We fully replicate the simpler Dana et al. paper, and broadly replicate the core findings for the other two projects, though with reduced effect sizes and a failure to replicate on some secondary measures. Finally, we attempt to connect behaviors to facilitate the understanding of how each fit within the broader literature. However, we find no connections across domains.

1. Introduction

Decision makers looking for an excuse to not give and instead make more self-interested choices has been the subject of an influential recent literature. We replicate three papers within this broad theme, each speaking to a different domain for the excuse not to give: (i) avoiding information in Dana et al. (2007); (ii) applying different risk postures in Exley (2015); and (iii) the motivated manipulation of a belief about the other party in Di Tella et al. (2015).[1] Our paper makes two important contributions. First, we examine the robustness of the original results, reproducing the experimental implementations and analyses in a well-powered study with a new draw from the data-generating process. Second, we explore connections between these three related phenomena across the different domains. As such, our paper provides a check on the robustness of the original results in each of the separate papers, while also providing evidence on how they fit together.

Our paper responds to the growing call for more-systematic reproductions of empirical research in economic journals (for example, see Maniadis et al., 2015, for experimental

Date: January 2022.

[1]Prominent other work on excuse seeking includes Linardi and McConnell (2011), where participants use excuses to lower time spent volunteering, Dana et al. (2006), where people are willing to engage in a costly, but quiet, exit to avoid participating in a dictator game, and Kajackaite (2015), where participants opt to remain ignorant about potential negative externalities from their actions.

replications).[2] Our paper offers results from a novel approach to this by integrating the design, implementation, and paper production into a graduate experimental-economics class project.[3] While the focus of our paper is on the research contribution from the replication exercise, we also stress the benefits of the process through the pedagogic spillovers from training students via replications.

Our findings mirror those of other replication collections: while we find qualitative evidence supporting the broad conclusions in each of the three studies, the effects are weaker (both statistically and economically) than the original for the majority of the papers we examine.[4] Moreover, for some of the secondary and supporting measures, our replications do not reproduce the original findings.

Our Exley (2015) replication mirrors the core result from the original study that participants use risk to a charity as an excuse to make more self-interested choices. However, we fail to replicate the finding that participants use self-risk (demonstrated by making themselves seem more risk-loving) as an excuse to donate less. Moreover, our statistical results are less significant than the original, primarily driven by smaller effect sizes.

For Di Tella et al. (2015), we replicate the finding that receiving an otherwise independent positive shock to the extent selfish behavior can (and does) benefit an individual leads to belief manipulation, with the individual distorting their belief to help justify a self-interested choice. However, while we replicate the original's finding about belief distortions over a matched partner's action, we do not detect distortions in beliefs about population-level actions. As such, the belief distortions we observe are more minimal. For the main result over the motivated beliefs in the partner we also find a decline in significance for the result. Where the result in the original has a confidence level in excess of 99 percent, this falls to 94 percent in our replication.

Finally, we strongly replicate the Dana et al. (2007) finding that decision-makers with the option to avoid information about the consequences of their choices on others act more selfishly. In this starker experiment, our quantitative effect is similar in size to

[2]Berry et al. finds that replications for papers published in the *American Economic Review* have been overall successful when a very broad view is taken on replicating a paper's main result, particularly for both highly cited work and experimental research. However, they tend to be embedded within broader extensions unless they fail to replicate. Given the importance of testing the robustness of scientific findings, Maniadis et al. (2015) and Coffman et al. (2017) offer a series of proposals to promote direct replications in economics, where this special issue is part of a clear (and hopefully ongoing) response.

[3]Better integrating replications within a pedagogic setting was one of the themes explored in 2017 American Economic Association conference sessions focused on replication issues in economics.

[4]Camerer et al. (2016) attempt to replicate 18 economics laboratory experiments published in the *American Economic Review* and *Quarterly Journal of Economics*. The authors replicate roughly two-thirds of the articles (depending on the measure of a successful replication), much lower than the expected 90% replication rate based on ex-ante power calculations. They also find substantially reduced effect magnitudes relative to original estimates.

that of the original, but with substantially increased confidence due to our study being overpowered.[5]

Our aim for these three replications, particularly given their provenance in an experimental class, is to go beyond just testing the robustness of the original results by also contributing to the understanding of them within the wider context of their literature. As such, the second arm of our study aims to connect excuse-seeking behaviors across the three domains: risk, information and beliefs.[6] Interestingly, we do not find substantial cross-context correlations for any of the excuse-seeking behaviors. In fact, if anything, they are negatively associated. We do find moderate associations for clearly selfish types across the studies, but little to connect an agent's generosity or excuse-seeking in one domain to how they act in another.

The most straightforward reading of the connections result is that excuse-seeking behaviors are orthogonal across domains for any fixed individual. However, we cannot rule out more-sophisticated stories; in particular that behavior may be non-stationary, creating a limit on the number of excuses an individual can make use of in a short amount of time to a fixed observer, such as to an experimenter. Though not identifiable in our data, some of these interesting loose ends will hopefully be explored further in subsequent work.

The remainder of this paper is organized as follows: Section 2 summarizes each of our three replications in turn: first outlining the original design; then listing any substantive changes made for the replication; followed by a comparison of the main results in the original and replication. Section 3 then briefly quantifies the connections across the three studies, before we conclude in Section 4.

2. Replication Results

2.1. **Overview.** Our replication was conducted physically at the Pittsburgh Experimental Economics Laboratory using an undergraduate population during the 2018-2019 academic year. We run two types of sessions: one with a compact version of Exley and the

[5]Sample sizes for the Exley and Di Tella et al. replications were selected for 90 percent power based on the original effect sizes/data. In contrast, to facilitate connections, the Dana et al. replication is run at the end of each session for both replications. The sample-size is therefore the sum of the other two, creating an over-powered sample.

[6]This extends the exercise Exley concludes with, where she connects her results to those in Dana et al..

Table 1. Sample Size for each Study

	Exley		Di Tella et al.		Dana et al.	
	Orig.	Repl.	Orig.	Repl.	Orig.	Repl.
Total participants	99	68	130	154	51	222
Effective participants	57	56	65	77	35	213

Note:
Effective sample sizes only include participants used in analysis. Exley drops censored and non-monotone subjects. The effective sample size for Di Tella et al. only lists Allocators in the core *modified* game. Dana et al. drops participants in the Hidden Information treatment without the Baseline underlying state. Sample sizes for the replications of Exley and Di Tella et al. are chosen using the original estimates to achieve 90% power and follow similar exclusion restrictions as the originals. Our Dana et al. study pools all participants across the two replication studies, including those not otherwise used for analysis. The only exception is 9 participants in the Di Tella et al. replication for whom we fail to record their Dana et al. decisions.

other with the core Di Tella et al. environment.[7] All sessions conclude with a within-subject variation of Dana et al..[8] We follow similar methods to Camerer et al. (2016), selecting sample sizes for the Exley (2015) and Di Tella et al. (2015) replications to achieve 90 percent power based on the original estimates. The Dana et al. (2007) replication uses pooled data from both replication samples. Table 1 lists the sample sizes for the original and replication studies.

2.2. **Exley Replication.**

2.2.1. *Review of Original Design.* Exley (2015) studies the use of risk as an excuse not to donate. An initial price-list elicits a point of indifference between a ten-dollar certain amount for the participant and a $\$X$ donation to a charity.[9] The \$10 self payment and its $\$X$ donation indifference point are then used as prizes in a series of price-lists across four comparisons:

$$\{\text{Self-risk } (P, \$10), \text{Charity-risk } (P, \$X)\} \times \{\text{Self-certain } \$0\text{–}\$10\ , \text{Charity-certain } \$0\text{–}\$X\}.$$

Each price-list offers a constant lottery on the left-hand side with a probability

$$P \in \left\{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\right\}$$

[7]We note that our studies are not pure reproductions of the originals. Changes include scaling payments to be in line with PEEL's standard practices and using z-Tree (Fischbacher, 2007) instead of either Qualtrics or paper and pencil. Finally, we read all instructions, available in Appendix D, aloud to ensure common knowledge amongst participants. We detail additional modifications for each study prior to discussing its findings.

[8]A simple symmetric-action game allowing for similar belief-distortion as those in Di Tella et al. is run before the Dana et al. decisions in the Exley sessions. This is discussed further in Appendix C.

[9]The maximum donation amount is \$30. Participants who never switch and therefore have indeterminate indifference points are classified as "censored" and dropped from the main analysis, following the procedure of the original study. The full distribution of $\$X$ is shown in Appendix A.

of winning the prize. The right-hand side contains certain amounts that vary over 20 evenly spaced increments, moving from \$0–\$10 in self-certain lists and from \$0–\$X in charity-certain lists. As such, we elicit two certainty equivalents for each type of lottery (self-risk and charity-risk): one measured in terms of a self-certain amount, the other in terms of a charity-certain amount.

The Exley valuations are used to assess whether risk to a charitable donation is used as an excuse not to give, thereby securing a higher payment for the self. This is assessed by examining the certainty equivalents when the elicitation has a *tradeoff* between the self and the charity (self-risk versus charity-certain amounts, and charity-risk versus self-certain amounts). Excuse-seeking behavior will therefore show up as a gap between the valuations of risky amounts to the charity measured in self-certain dollars relative to when they are measured in charity-certain dollars. Controlling for the certainty equivalents when there is no tradeoff between the benefactors of the lottery and certain amounts (self-risk versus self-certain amounts, and charity-risk versus charity-certain amounts) allows for distinct risk preferences for the charity and the self.

The core finding in Exley is that participants become significantly more risk-averse when evaluating a risky donation to charity against a self-certain amount. This result holds at $P = 0.95$, demonstrating that even small risks to the charity are used as an excuse to reduce donations. Similarly, participants become more risk-seeking when evaluating self-lotteries in terms of charity-certain amounts.[10]

2.2.2. *Replication Design Changes.* The most significant changes we make are: (i) changing the charity to the University of Pittsburgh Medical Center Children's Hospital from the American Red Cross; (ii) reducing the number of separate price lists by reducing the number of probabilities; and (iii) using a switch-point elicitation method rather than binary comparisons in each of the twenty choices.

Switching the charity to a local Children's Hospital tailors the elicitation to a cause familiar to participants. Moreover, we read aloud a charity plea during the instructions and add a shortened version to each price list with a charity component to increase the salience of the donation.

The remaining changes lower the number of decisions in the study, improving the flow and reducing cognitive load. We elicit lottery valuations only for $P \in \{0.05, 0.25, 0.75, 0.95\}$, reducing the number of price lists in the study from 30 to 17.[11] Finally, in the original

[10]At the end of the study, participants complete a within-subject implementation of Dana et al. (2007). Exley classifies subjects based on their decisions and finds a correlation between excuse-seeking behaviors. We return to this result in our discussion of connections in Section 3.

[11]We remove the 12 elicitations with $P \in \{0.1, 0.5, 0.9\}$ and an additional price-list between charity-certain amounts and a \$5 certain payment to self.

Table 2. Core Replication Results: Exley (2015)

Variable	Original		Replication	
	Coef.	p-value	Coef.	p-value
Charity-risk	0.06	0.946	0.53	0.661
Tradeoff	5.30	**0.011**	2.14	0.438
Charity-risk × *Tradeoff*	−15.09	< **0.001**	−6.82	**0.081**
Charity-Tradeoff effect	−9.79	< **0.001**	−4.68	**0.031**

Note: OLS estimates from Exley (2015) Table 2 (the original columns) and our replication (the replication columns). The outcome variable is the relative valuation for the lottery, with indicators for the lottery being for charity, the presence of a self-charity tradeoff (differing benefactors for the lottery and the varying certain amounts), and their interaction. All p-values derived from participant clustered tests. In the *Charity-Tradeoff* effect row (not given in the original study tables) we provide the sum of *Tradeoff* and interaction terms. Full regression tables are available in Appendix A.

study participants were required to separately click on one of the two options for each of the twenty increments in the price list, while we instead use a switch-point design.[12] These changes therefore lower the number of distinct decisions from 600 in the original to 17 in ours.

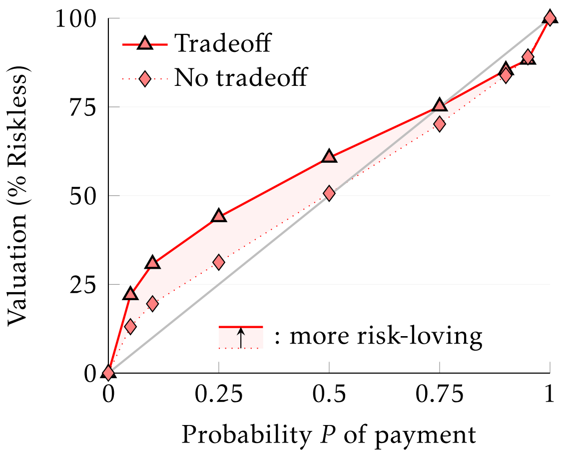
2.2.3. *Results.* Our main results for the Exley replication are: (i) a qualitative replication of the result that participants use risk to a charity as an excuse to donate less, though with an attenuation of the effect size; and (ii) a failure to replicate the secondary result that participants similarly use risk to self as an excuse to donate less.[13]

In Figure 1, we graph the average relative valuation for each separate probability P. The elicited switch-point is expressed as a percentage of the maximum certain payment—\$10 for self-certain amounts or \$X for charity-certain amounts. As such, a risk-neutral agent's valuation increases linearly with the probability P, shown by the grey diagonal line in each figure. A common behavioral response is to instead over-value low probability lotteries and under-value high probability ones: the classic inverse-'S' shape. We find evidence for probability weighting for both self- and charity-risk.

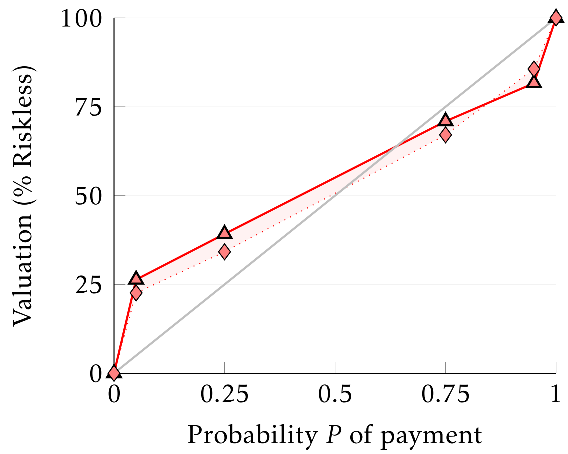
Figure 1 reports the average results from the original Exley study in the left three panels and from our replication study in the right three. The three rows in the figure each help illustrate a core result. The figure in the top row compares relative valuations at every probability for self-risk lotteries measured with and without a tradeoff (so, against

[12]We therefore enforce a single switch-point (monotonicity). However, most participants in the original Exley study demonstrated monotone preferences already.

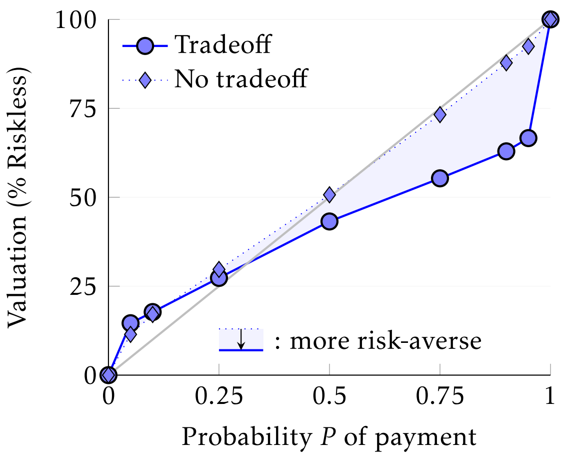
[13]We also find lower censoring in our data, defined as never switching to the charitable donation in the first price-list. We attribute this to participants potentially having greater valence with the changed charity. Per the core Exley analysis, our main results are presented without the censored participants, but are robust to including them; see Appendix A.



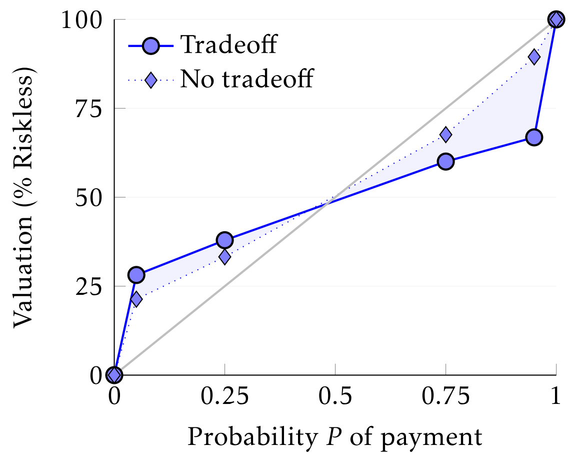
(A) Exley (Self-risk)



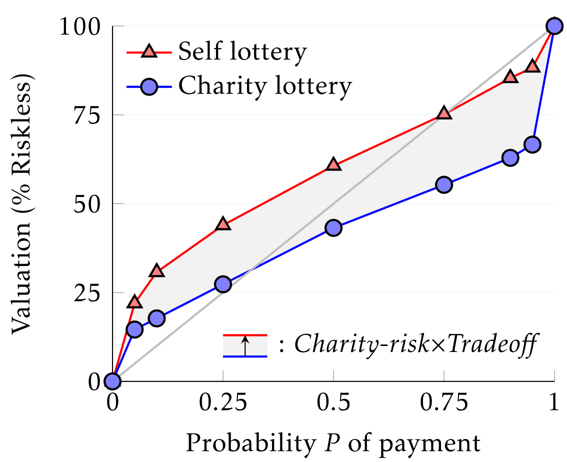
(B) Replication (Self-risk)



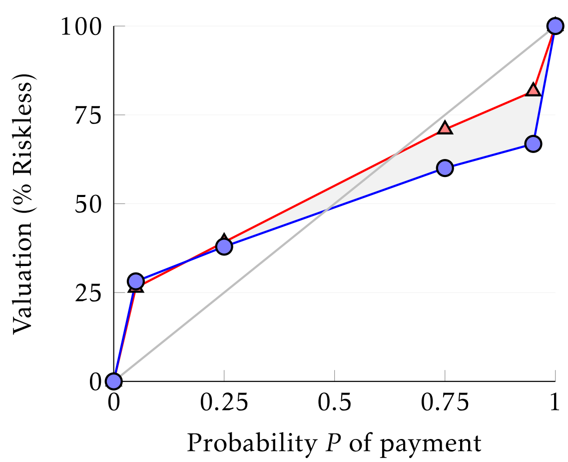
(C) Exley (Charity-risk)



(D) Replication (Charity-risk)



(E) Exley (Tradeoff)



(F) Replication (Tradeoff)

Figure 1. Self-Tradeoff Lists: Relative Valuations

charity-certain and self-certain amounts, respectively). The middle panels do the same with the relative valuations for lotteries with charity-risk. Finally, the bottom two panels fix the presence of a self/charity tradeoff, allowing for a comparison of relative valuations for self-risk and charity-risk lotteries.

Table 2 complements the average results in the figure by reporting the regression estimates from the core inferential regression, where data is pooled across all different lottery probabilities. The relative valuation for each lottery is regressed on three dummy variables: (i) the elicitation being over a lottery benefiting the charity (*Charity-risk*); (ii) the elicitation having a tradeoff between the lottery and the certain payment (*Tradeoff*); and (iii) the interaction of the two (*Charity-risk* × *Tradeoff*), for elicitations where charity-risk is evaluated in self-certain amounts.[14]

The *Charity-Risk* coefficient speaks to whether the average risk preference for lotteries with charitable donation prizes is distinct from those for lotteries with self-payment prizes when evaluated in terms of certain amounts in the same domain. Here we fully replicate the economically negligible and insignificant estimated effect found by Exley. This establishes that participants in both studies have similar risk preferences for charity- and self-risk in the absence of a tradeoff, allowing us to make inferences about how the presence of a tradeoff impacts risk preferences.

The next pair of coefficients in Table 2 examine the relative pricing effect of a tradeoff between self and charity in the evaluation. In the original study, Exley finds a significant positive effect on the *Tradeoff* term and a significant negative effect on the *Charity-Risk* × *Tradeoff*) interaction. The quantitative interpretation of the original estimates is that there is a 5.3 percentage-point increase in the average valuation for lotteries with self-risk when measured in terms of a charitable donation. This effect is shown in Figure 1(A) as the shaded gap between the two lines. In contrast, there is a 9.8 percentage-point decrease in the average valuations when assessing lotteries with charity-risk in terms of self-certain payments. This effect can be seen across the different probabilities in Figure 1(C), where the average effect is calculated by adding together the *Tradeoff* and *Charity-Risk* × *Tradeoff* coefficients in Table 2.[15] Finally, considered on its own, the -15.1 percentage point estimate for the *Charity-Risk* × *Tradeoff* interaction in Exley indicates the average difference between self-risk and charity-risk lotteries measured with a tradeoff. This difference is illustrated as the gap between the two lines in Figure 1(E).

We only partially replicate the excuse-seeking results, shown by comparing the findings of the original to those of the replication in Table 2 and Figure 1. First, we do not

[14]Paralleling Exley's analysis, in Appendix A we provide a full replication of Table 2. The table includes interval and Tobit estimates and a specification including dummies for the different probabilities.

[15]The last row of Table 2 indicates this net effect using a linear combination.

replicate the original findings for the *Tradeoff* coefficient. Our estimate is in the same direction as the original but with a much smaller effect (2.1 percentage points).[16] Comparing across probabilities in Figure 1(B), the reduced effect sizes hold even for the low-probability lotteries for our replication, whereas Exley finds a larger effect.[17] As such, we do not replicate the finding that the participants become more risk-loving over self-risk lotteries when measured in terms of a donation.

We replicate the qualitative finding that the participants become more risk-averse when assessing charity-risk lotteries. Looking at the estimates from Table 2, our replication finds a marginally significant effect of -6.8 percentage points on the *Charity-Risk* × *Tradeoff* interaction. This is illustrated, split by probability, in Figure 1(F). The last row of the table takes into account the correlation between the two tradeoff estimates. We find that their sum, -4.7 percentage points, is smaller but more tightly estimated. The conclusion is that we do find a significant reduction ($p = 0.031$) for the charity-risk valuations with a tradeoff. Figure 1(D) illustrates that, paralleling the original study, the differences in the treatment of charity-risk is driven by lotteries with a high probability of winning.

Result 1 (Exley replication)**.** *We partially replicate the Exley (2015) result but with the effects attenuated by over a half. While we do not replicate the finding that participants become more risk-loving for themselves as an excuse not to give, we do find that they use risk to a charity as an excuse to give less.*

2.3. **Di Tella et al. Replication.**

2.3.1. *Review of Original Design.* Di Tella et al. (2015) focuses on the use of motivated beliefs over others' actions as an excuse to take from them. At the start of the experiment, participants who act as *Allocators* are randomly and evenly divided into two types that differ only in their ability to redistribute 20 tokens between themselves and a randomly matched partner from the initial equally split endowment. One type (Able = 2) can shift at most two tokens per account while the other (Able = 8) is given greater agency and allowed to shift at most eight.

Unlike a straightforward dictator game, the Allocator's partner is given the option to receive a side-payment. Doing so, however, lowers the monetary value of the tokens from \$1.50 to \$0.50 each. Without any knowledge of their matched Allocator's type or re-distribution decision, the partner must either forego the side-payment so the pair can

[16]The standard-error of our *Tradeoff* coefficient is almost identical to that in Exley despite the estimated effect being halved. This means the reduced significance is not driven by greater variability in our sample.
[17]We present the main specification for both samples split by probability in Appendix A.

have high-value tokens or make a ‘corrupt’ decision to accept the additional $5 for themselves in exchange for the pair having low-value tokens.[18]

Each pair begins with an initial $(10,10)$ token-endowment.[19] Without knowing whether their matched partner chose to take the side-payment, Allocators decide how many tokens to move across the accounts. Allocators who are confident that their partner will pick the corrupt decision—taking the side-payment in exchange for lowering the value of the tokens—may feel justified in moving tokens from their partner’s account to their own. The core idea is that participants who want to act selfishly will alter their beliefs over their partner’s corruption, justifying their redistribution. This idea is assessed through the variation in Allocator types (Able = 8 or Able = 2). Each Allocator’s actual redistributional ability is independently assigned, and it is common-knowledge that their partner does not know their type when making the token-value decision. As such, the Allocators’ maximum benefit from redistribution shifts across treatment, but the corruption level of their partners remains constant.

The independence between actual corruption and re-distributional ability allows us to compare Allocator beliefs by type to measure motivated reasoning. To this end, two levels of the Allocators’ beliefs are elicited: one over the corruptness of their matched partner and one over the session-level proportion of corruption.[20]

The core finding in Di Tella et al. (2015) is that the Able = 8 Allocators are much more pessimistic about their partners’ action than the Able = 2 types. That is, an independent shift in the ability to benefit from redistribution substantially alters Allocators’ beliefs in a direction that justifies self-serving behavior, identifying motivated beliefs. The study finds a significant response for both beliefs about the matched partner’s corruption and beliefs over the session-level proportion of corrupt actions.

2.3.2. *Replication Design Changes.* Our main change is updating the instructions for Allocators to *explicitly* state that their partner is unaware of their type (Able= 2 or Able= 8). This information is only indirectly provided in the original experiment. Our change therefore removes any ambiguity over this core design feature.

[18]This describes the *Modified* game from Di Tella et al. (2015) in which the corrupt option is only the dominate action if the Allocator takes over four tokens. This uncertainty over corruptness allows for a clearer identification of motivated beliefs than the *Basic* game, also included in the original study, where the dominant response is always the corrupt action.

[19]Tokens are earned in a real-effort task preceding the main decisions.

[20]Allocators receive $1 for correctly predicting their partner’s action and $5 for accurate session-level beliefs.

Table 3. Core Replication Results: Able= 8 effect in Di Tella et al. (2015)

Dependent Variable	Original		Replication	
	Coef.	*p*-value	Coef.	*p*-value
Tokens Taken	5.150	< **0.001**	2.670	< **0.01**
Is Corrupt	0.325	**0.009**	0.225	**0.056**
%-Corrupt	0.151	**0.008**	0.030	0.558

Note: OLS estimates from Di Tella et al. (2015) Table 3 (the original columns) and our replication (the replication columns). Coefficients show the result of regressing an indicator variable for being an Able= 8 Allocator for three different dependent variables: (i) The number of tokens redistributed, positive if taken from their partner and negative if given; (ii) an indicator variable for if they belief their partner took the corrupt action; and (iii) their belief over the session-level proportion of corruption. All *p*-values derived from participants clustered tests. Full regression tables are available in Appendix A.

Additionally, allocation decisions in our replication happen prior to the belief elicitations, preventing anchoring effects. Finally, we convert all payments from Argentinian Pesos to USD.[21,22]

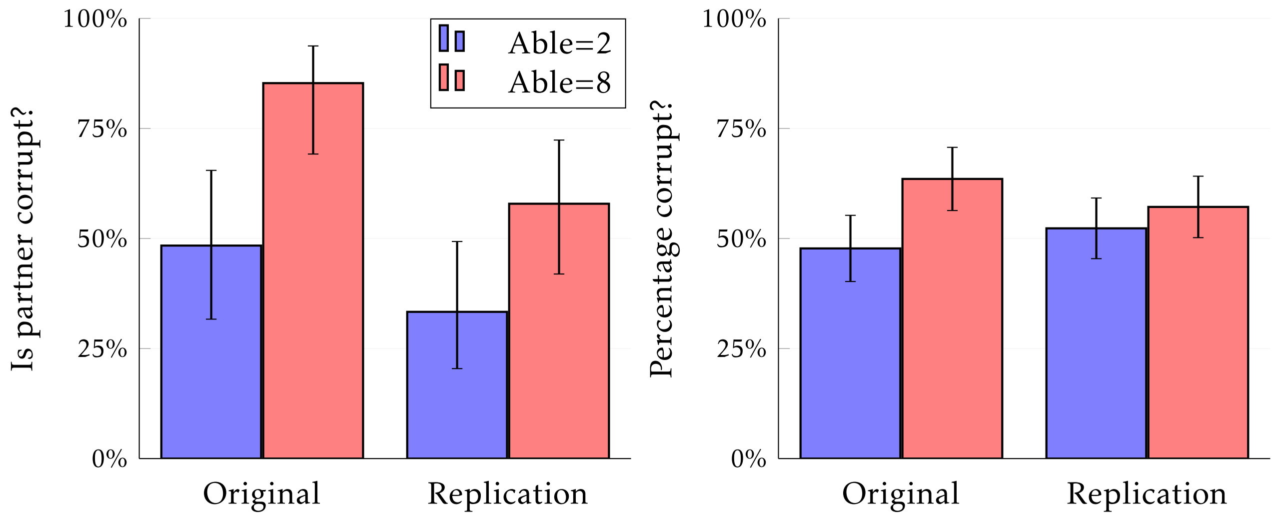
2.3.3. *Results.* Our core results for the Di Tella et al. replication are: (i) a replication of the finding that Allocators who gain more from redistribution are more likely to distort their beliefs over their partner's corruption, though with a smaller effect size; and (ii) a failure to replicate motivated beliefs about session-level corruption.

Figure 2 and Table 3 outline the core results from both the original study (the left panel/columns) and our replication (the right panel/columns). The table reports treatment-effects for Able= 8-Allocators, using the Able= 2-Allocators as the control. Each estimate comes from a separate regression, following Di Tella et al.'s specifications.[23] The dependent variables for the three regressions are: (i) the number of *Tokens Taken*, measuring the degree to which Allocators redistribute; (ii) the elicited belief about their partner choosing the corrupt option (*Is Corrupt*), measuring partner-level beliefs; and (iii) the elicited belief about the proportion of partner-participants in the session who take the corrupt option (*%-Corrupt*). Complementing the last two rows in the Table, Figures 2a and 2b illustrate the average elicited belief by Allocator type. Partner-level beliefs are shown in the left panel, and session-level on the right.

[21]Our payments are converted 1:1 from the original, rather than using an exchange rate. This means the same *numbers* are used in both studies.

[22]We also change the payment rule for the session-level elicitation. We give a smaller reward for beliefs 'close' to the true proportion, smoothing out the marginal incentives at extreme beliefs.

[23]Full regression results are reported in Appendix A.



(A) Is corrupt?

(B) Percent corrupt

FIGURE 2. Replication results for Di Tella et al. (2015)

Note: Confidence intervals (95 percent coverage) indicated for: predicted proportion in panel (A), and predicted level in panel (B).

The first result in the table indicates that Able= 8 Allocators take significantly more tokens in both studies. However, our effect size is much smaller than the original.[24] The second result is the key finding: giving Allocators a greater ability to redistribute moves their beliefs in a self-serving direction. This is illustrated in the second row of Table 3 and Figure 2(A). The treatment effect in the original study is large, with beliefs of Allocators in Able= 8 shifting 32.5 percentage points relative to those in Able= 2. Our replication finds a similar but smaller effect. The difference is reduced to 22.5 percentage points, lowering the significance level to a more-marginal p-value of 0.056.

Finally, the last row of Table 3 and Figure 2(B) examine session-level beliefs, where we fail to replicate the original finding. Being an Able= 8 Allocator in the original study is associated with a 15.1 percentage points increase in the belief over what proportion of the population acted corruptly. In contrast, we find only a negligible difference of 3 percentage points.

We note two potential factors for the failure to replicate the session-level belief distortion. First, Allocators may wish to engage in the minimum necessary belief distortion needed to justify their selfish token redistribution. Allocators are only directly affected by the decision of their partner, so they may only need to “justify" their action in that

[24]Although we do not focus on partner actions, as they are independent of Allocator type and therefore do not affect inference over beliefs, we note that we observe significantly less corruption in the replication sample. While 66 percent of partners were corrupt in the original, only 26 percent are in our replication.

respect. Second, our change to the instructions makes it very clear that partners are unaware of their matched Allocator's type. Although this information is disclosed in the original study, we increase its saliency which may make it more difficult for a participant to act *as if* they did not know it. Ging-Jehli et al. (2019) also replicate Di Tella et al. using the exact *Modified* game and instructions from the original Di Tella et al. (2015). They find qualitatively similar results to the original paper for both types of beliefs, adding some credence to this small modification driving the difference.

Result 2 (Di Tella et al. replication)**.** *We partially replicate the Di Tella et al. (2015) result but with attenuated effect sizes. While we do find that an increased ability to re-distribute moves partner-level beliefs in a motivated direction, we do not find any effect on session-level beliefs.*

2.4. **Dana et al. Replication.**

2.4.1. *Review of Original Design.* Dana et al. (2007) explores information avoidance as an excuse for selfish behavior. Participants make binary choices in a modified dictator game between two (self,other) payment bundles, where the 'other' is another participant in the session. A between-subject design is used to examine how access to "moral wiggle room" can increase selfish actions.

In the *Baseline* treatment, participants decide between payment bundles of ($5, $5) and ($6, $1). The selfish choice here is the ($6, $1) bundle, which maximizes the self-payoff. Participants can alternatively act altruistically and give up $1 for themselves to increase the other's payment by $4, thus achieving perfect equity. In the *Hidden Information* (HI) treatment, participants similarly make choices over two bundles, however, an unknown state variable $Y \in \{\text{Unaligned}: \$0, \text{Aligned}: \$4\}$ is incorporated into the the other participants payoff. The dictator therefore must choose between the bundles ($\$6, \$1 + Y$) and ($\$5, \$5 - Y$). When the game's hidden state is *Unaligned*, the bundles are identical to the Baseline treatment. However, when the state is instead *Aligned*, the choice is between the bundles ($6, $5) and ($5, $1) and the same bundle maximizes payoffs for both the self and other. Knowing Y is therefore key for the dictator to understand how giving up $1 impacts the other's payoff. The state's realization can be *costlessly* revealed to the dictator before they make their decision by simply clicking a button on their screen. However, choosing not to know allows dictators to make the decision to keep an extra $1 for their self without knowing whether this choice hurts or helps their partner.

The key finding from Dana et al. is that dictators in the HI treatment avoid costless information about Y, instead taking $6 for themselves without knowing its impact on the other's payoff. Dictators in the Baseline treatment, on the other hand, mostly give up a dollar for themselves to ensure the fair ($5, $5) allocation. The proportional increase

Table 4. Core Replication Results: Dana et al. (2007)

Variable	Original		Replication	
	Coef.	p-value	Coef.	p-value
Baseline (\$6, \$1) choices	0.266	–	0.174	–
HI (\$6, \$1) choices (at $Y = 0$)	0.625	**0.044**	0.551	< **0.001**
Difference	0.361	**0.033**	0.377	< **0.001**

Note: Estimates are reported for both Dana et al. (2007) (the original columns) and our replication (the replication columns). Coefficients show the results of two-sided Fisher's Exact test comparing the difference in proportions picking (6, 1) when $Y = 0$ in the Hidden Information treatment to the Baseline treatment and the OLS results of regressing being in the HI treatment when $Y = 0$ on picking (6, 1). Results from the original are calculated using information from Tables 1 and Tables 2 in Dana et al. (2007). Participant choices are reported in Appendix A.

in dictators choosing (\$6, \$1) in the HI treatment when $Y = 0$ over the Baseline treatment where dictators face the same bundles with full information imposed, drives the conclusion that participants use moral wiggle room. That they use ignorance about the externalities of their choice as an excuse to engage in self-serving behavior.

2.4.2. *Replication Design Changes.* Our design modifies the Dana et al. design to be within-subject, closely following the Exley (2015) implementation. This differs from the original in two key ways. First, we modify the recipient in the 'other' role from another participant, such as in the original study, to a charity.[25] Second, the within-subject design forces the participant to make three decisions. In one, they make the baseline decision with $Y = \$0$ between (\$6, \$1) and (\$5, \$5). In another, $Y = \$4$ and thus their decision is between (\$6, \$5) and (\$5, \$1), where the selfish and altruistic choices coincide. And finally, one choice is made with hidden information (randomized at the participant level here, and at the session-level in the original) where the participant can costlessly click a button to reveal exactly what charitable donation each bundle has.[26]

2.4.3. *Results.* We strongly replicate the core Dana et al. (2007) result that participants purposefully avoid information as an excuse to act selfishly. We now detail the levels and inferential conclusions from the original and our replication.

Table 4 and Figure 3 show the proportion of participants choosing the selfish (\$6, \$1) bundle in the Baseline treatment and in the HI treatment when $Y = 0$.[27] In the original paper, only 27 percent of subjects made the selfish choice under full information. However, for an identical choice placed behind a small informational obstacle, the proportion

[25] We again use the local Children's hospital, whereas the Exley paper uses the American Red Cross.

[26] Our only deviation from Exley is to implement one decision from two randomly selected participants per session instead of paying one decision per participant.

[27] The exclusion of participants in the HI treatment for whom $Y = 4$ follows the original Dana et al. analysis.

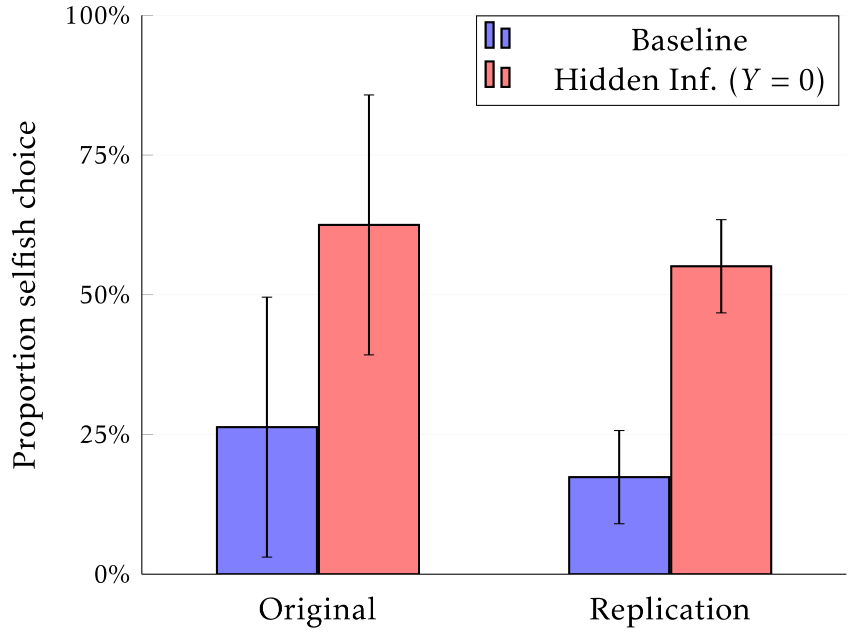


Figure 3. Replication results for Dana et al.

making the selfish choice jumps up to 63 percent. As such, approximately 36 percent of the subjects can be said to use information avoidance as an excuse. We find a remarkably similar proportion of excuse-seeking behavior. In our Baseline treatment, only 17 percent opt for the selfish option.[28] This jumps up to 55 percent in the analogous Hidden Information treatment decision. This 38 percentage point gap indicates a quantitatively large number of moral wigglers, but is similar in size to the original estimate.[29]

The inferential p-values given in Table 4 are much smaller in our replication than in the original study. However, a large part of this is due to our much larger sample. Unlike the Exley and Di Tella et al. replications, where sample sizes were chosen to achieve 90 percent power, our Dana et al. replication sample size is the sum of the total number of participants in the other two, giving it much greater power.[30] To remove this much greater statistical power, we create 1,000 sub-samples from our data at an identical size to the original study by randomly selecting 19 baseline treatment observations and 16 HI treatment with $Y = 0$. We find an average p-value of $p = 0.08$ when running the same

[28]This slight difference in levels could result from an imprecise estimate in the original, given the smaller sample size. However, it may also be morally harder to take \$4 from a Children's Hospital play room to make yourself \$1 richer.

[29]Following the original study's inference, we use a χ^2-test to compare the difference in proportions. In the original study, this yields a p-value of 0.044. In contrast, for our replication this test yields a $p < 0.001$.

[30]The Dana et al. sample is larger than the sum of the effective samples for the other two, as it includes censored participants from Exley and partner participants from Di Tella et al.

inferential χ^2-test as the original. We therefore attribute the increase in inferential power in our replication to our over-powered sample.

Result 3 (Dana et al. replication)**.** *We replicate the original findings with similar effect sizes to the original study. Our replication yields a much smaller p-value, primarily driven by greater statistical power from a much larger sample.*

2.5. **Replication Summary.** We qualitatively replicate the main results from three separate studies to look at how excuses in the form of risk, distorted beliefs, and information avoidance can impact selfish behavior. For two of the three studies, the effect size in the replication is diminished both in terms of magnitude and significance level. Moreover, for secondary analysis in these two studies, we replicate in direction only.

We also find less selfish behavior overall in our sample. In the Exley (2015) replication, we have fewer participants who refuse to give up \$10 for themselves for even a tripled donation to charity. In the Di Tella et al. (2015) replication, fewer participants select the corrupt option and Allocators in general take fewer tokens. Finally, fewer participants take the selfish option in State 1 in both the baseline and HI treatments in our Dana et al. (2007) replication.[31]

3. Connections

We conclude by exploring connections across the three types of excuse-seeking behavior. We attempt to both replicate and extend the connections made in Exley (2015, Table 4) between the use of risk and the use of information avoidance as excuses not to give.[32] To extend this idea further, we also try to connect the motivated beliefs used as an excuse to act selfishly in Di Tella et al. (2015).[33]

Exley examines connections across the excuse-seeking domains through an interaction of Dana et al. types with the core three regressors (cf. Table 2). However, we do not find any interaction between the use of information-avoidance to the use of risk as an excuse in this analysis for our replication. We now turn to a simplified comparison over estimated types to set this exercise in a unified context and establish connections to motivated beliefs as an excuse.

For each study we create a parallel typology to that identified in Dana et al.: those who act selfishly (even without an excuse), those who make use of a given excuse to justify

[31] However, a similar proportion of participants engage in moral wiggle, making the difference between the two similar to the original study.

[32] We reproduce Exley Table 4 in Appendix B using both the original and our replication samples.

[33] We focus on connections between Di Tella et al. and Dana et al.. We did add a module to our Exley which modifies and simplifies the Di Tella et al. environment, allowing each participant to act as both the Allocator and the partner. We find no evidence of motivated beliefs in this module, thus making the data not useful for connecting behavior. We therefore relegate it to Appendix C.

acting selfishly, and those that act generously (and do not use excuses). We drop observations of participants who cannot be classified as one of the three mutually exclusive Dana et al. categories and then attempt to connect types across the different experiments.

The stark nature of the choices in Dana et al. allow the three types to be pinned down with little latitude in interpretation.[34] In the other two settings, the *behavior* associated with each type is easily identifiable, but the decisions also allow for an intensity: the valuations for charitable donations in Exley and the number of tokens redistributed in Di Tella et al.. We approach the decision over threshold intensities for each type by matching the in-sample marginal distributions of the Dana et al. assessments. Our approach therefore takes the intensive margin decision out of our hands by ensuring we have the same proportion of selfish/excuse-seeking/generous types in each measure for each sample. We examine the connections by looking at the joint distributions and the extent they overlap for each type.[35]

Using the type pairs defined by decisions in the Dana et al. study on one side and those defined by parallel behaviors in either Exley or Di Tella et al. on the other, we provide estimates of the pairwise correlation in Table 5 alongside the p-values for tests of significance.

The first three data rows report results using the original Exley data. Our results suggest no relationship across excuse-seeking types. However, we find significant positive correlations for the *Generous* and *Selfish* types across contexts (as well as strong negative correlations between them). The next three rows present results using data from our Exley replication. We similarly find a significant positive correlation across the two *Selfish* types. While there is some evidence for negative correlations across types (in particular, the *Selfish* and *Excuse-seeking* classifications), we fail to connect either the *Excuse-seeking* or *Generous* classifications across contexts. Finally, in the last three rows, we repeat the exercise for our Di Tella et al. replication. While this is the smallest sample of the three

[34]Selfish individuals always maximize their own payoff, while excuse-seekers only do so in the presence of moral wiggle-room. Finally, generous individuals seek full information to make the altruistic choice. The full classification, which follows the one used by Exley, is presented in Appendix B.

[35]In the Exley studies (both the original and replication), we define two cutoffs: one over the initial normalization amount X (inversely related to preferences for the charity) and one over the under-evaluation of the charity lottery in self-certain amounts compared to charity-certain amounts, given the similarity of self- and charity- risk preferences, at $P = 0.95$ (directly related to excuse seeking). In the Di Tella et al. study we restrict attention to Able = 8 Allocators—those who had access to an excuse—and define two cutoffs over the number of tokens transferred: one when they believe their partner is corrupt, and one when they do not. For our Exley replication, the marginal distribution over the (Excuse-seeking, Generous, Selfish) types is therefore (36,18,12) using risk decisions and (36,19,11) using Dana et al. information decisions, while the original study has the marginal distributions of (16,40,42) for risk and (20,36,43) for information. Our Di Tella et al. replication has a marginal-type distribution of (12,15,8) for belief decisions and (13,16,6) for information ones. Full details of the type classifications for all studies can be found in Appendix B.

Table 5. Pairwise type correlations with Dana et al. (2007) typology

Sample	Type in sample	Type in Dana et al.					
		Excuse-Seek.		Generous		Selfish	
		Coef.	*p*-Value	Coef.	*p*-Value	Coef.	*p*-Value
Original Exley	Excuse-seek.	-0.018	0.859	0.132	0.196	-0.112	0.270
	Generous	0.146	0.151	0.291	**0.004**	-0.400	< **0.001**
	Selfish	-0.132	0.197	-0.387	< **0.001**	0.481	< **0.001**
Replication Exley	Excuse-seek.	0.083	0.506	0.110	0.379	-0.245	**0.047**
	Generous	0.217	**0.080**	-0.089	0.478	-0.183	0.142
	Selfish	-0.359	**0.003**	-0.039	0.753	0.527	< **0.001**
Replication Di Tella et al.	Excuse-seek.	-0.306	**0.074**	0.062	0.723	0.310	**0.070**
	Generous	0.290	**0.091**	-0.099	0.570	-0.241	0.164
	Selfish	0.004	0.982	0.047	0.789	-0.067	0.702

Note: This table reports correlation of Excuse, generous, and selfish behavior in the Dana et al. decisions and the respective definition of that behavior in both Exley studies and the Di Tella et al. replication.

comparisons (as we can only use Able = 8 Allocators with a clear Dana et al. type) we do find some significant results. However, they move in the opposite direction from what we would expect. Excuse-seeking types in the Dana et al. classification—those who use information-avoidance as an excuse—are significantly less likely to be an Allocator with a pessimistic view of their partner's corruptness who consequently redistributes tokens in their own favor. Instead, they are more likely to be generous with their token redistribution. In contrast, the excuse-seeking types in the Di Tella et al. replication—those who distort their beliefs—are more likely to be classified selfish, even without an excuse, in their Dana et al. decisions. Indeed, for the Di Tella et al. replication we do not find any of the three types to have a significantly positive relationship with the parallel definition.

Although we cannot test any hypothesis over mechanisms here, we propose two potential explanations for the broad lack of connections outside of the selfish types. First, the lack of connections between excuse-seeking behaviors may be derived from participants offsetting an excuse used in a prior setting with more-moral behavior in the next, and vice versa. This explanation has similarities to moral licensing, especially when the costs to consistent excuse-seeking differ across contexts see Gneezy et al., 2012). Second, the inconsistency in behaviors across choice environments might simply reflect the sensitivity of results to measurement methods. The impact of measurement error in experiments has been documented in other domains, including risk preferences for example Gillen et al., 2019; Friedman et al., 2014; Schildberg-Hörisch, 2018).[36]

Result 4 (Connections across Behaviors)**.** *We do not find significant positive associations between the excuse-seeking behaviors in our three distinct environments—in fact, the only meaningful relationship is negative. However, we do replicate the stronger correlations of the starker* Selfish *classifications in the Exley replication.*

4. Conclusion

Our paper presents results from three replications which were planned and implemented within a graduate class, integrating both a research and pedagogic component into a project answering the call for greater replication in economics. Our main findings replicate the core results in Exley (2015), Di Tella et al. (2015) and Dana et al. (2007), albeit with smaller effect sizes for the first two. The results are therefore similar to the quantitative attenuation described in the replications in Camerer et al. (2016).

We fully replicate the simpler Dana et al. study, reproducing similar quantitative effects albeit with increased significance due to a larger sample. For the Exley and Di Tella et al. replications, which look across more measures, we fail to replicate some important secondary analyses in each. For Exley we do not reproduce the finding that participants act more risk-seeking for self-lotteries priced in terms of a charitable donation. In Di Tella et al., the belief distortions we uncover in our replication are more minimal, focused more precisely on just the specific partner rather than the overall population.

In interpreting the parts we fail to replicate, it is important to take into consideration the tertiary changes within the implementations we make—the identity of the charity, the positioning of the belief elicitation, the specific population we experiment on, etc. While one interpretation of our results is as a check on the reproducibility of the original, another is as a sensitivity check on the robustness of the result to slight changes.

In the final section of our paper, we move beyond a pure replication by attempting to connect different excuse-seeking behaviors across domains. However, here we find a puzzling result. Fixing the decision maker, some of the more-clearly selfish behaviors are predictive across domains but we do not find substantial positive correlations over the use of excuses (nor generosity). Further research here is warranted, particularly with regard to understanding the repeatability of excuses; on the degree to which conscience accounting or other non-stationarities may inhibit our ability to detect such connections.

[36]For example, while our Di Tella et al. replication has the qualitatively weakest connections across the classified types, it also has the largest variation across the types of measurement: not only do the two measures differ in one being strategic and the other a decision, the nature of the impacted 'other' also changes (another participant or a charity).

References

Berry, James, Lucas Coffman, Rania Gihleb, Douglas Hanley, and Alistair Wilson, "Assessing the rate of replication in economics," *American Economic Review*, 2017, *107* (5), 27–31.

Camerer, Colin F, Anna Dreber, Eskill Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu, "Evaluating replicability of laboratory experiments in economics," *Science*, 2016.

Coffman, Lucas, Muriel Niederle, and Alistair Wilson, "A proposal to organize and promote replications," *American Economic Review*, 2017, *107* (5), 41–45.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes, "What you don't know won't hurt me: Costly (but quiet) exit in dictator games," *Organizational Behavior and Human Decision Processes*, 2006, *100* (2), 193–201.

______ **, Roberto A Weber, and Jason Xi Kuang**, "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 2007, *33* (1), 67–80.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *Review of Economic Studies*, 2015, *83* (2), 587–628.

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental economics*, 2007, *10* (2), 171–178.

Friedman, Daniel, R. Mark Isaac, Duncan James, and Shyam Sunder, *Risky Curves: On the Empirical Failure of Expected Utility*, New York: Routledge, 2014.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with measurement error: Techniques with applications to the Caltech Cohort Study," *Journal of Political Economy*, 2019, *127* (4), 1826–1863.

Ging-Jehli, Nadja R, Florian Schneider, and Roberto A Weber, "On self-serving strategic beliefs," February 2019. University of Zurich, Department of Economics, Working Paper.

Gneezy, Ayelet, Alex Imas, Amber Brown, Leif D. Nelson, and Michael I. Norton, "Paying to be nice: Consistency and costly prosocial bheavior," *Management Science*, 2012, *58* (1), 179–187.

Kajackaite, Agne, "If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment," *Journal of Economic Behavior & Organization*, 2015, *116* (C), 518–524.

Linardi, Sera and Margaret McConnell, "No excuses for good behavior: Volunteering and the social environment," *Journal of Public Economics*, 2011, *95* (5), 445–454.

Maniadis, Zacharias, Fabio Tufano, and John A List, "How to make experimental economics research more reproducible: Lessons from other disciplines and a new proposal," in "Replication in experimental economics," Emerald Group Publishing Limited, 2015, pp. 215–230.

Schildberg-Hörisch, Hannah, "Are risk preferences stable?," *Journal of Economic Perspectives*, 2018, *32* (2), 135–154.

Tella, Rafael Di, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman, "Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism," *American Economic Review*, 2015, *105* (11), 3416–42.