INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Social Risk: the Role of Warmth and Competence

Jeffrey V. Butler and Joshua B. Miller

Working Paper n. 522

This Version: July 19th, 2014

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Social Risk: the Role of Warmth and Competence

Jeffrey V. Butler[a]
EIEF

Joshua B. Miller[b] *
Bocconi University and IGIER

This version: *July 19, 2014*

Abstract

Previous research has documented a behavioral distinction between "social risk" and financial risk. For example, individuals tend to demand a premium on the objective probability of a favorable outcome when that outcome is determined by a human being instead of a randomizing device (Bohnet, Greig, Herrmann, and Zeckhauser 2008; Bohnet and Zeckhauser 2004). In this paper we ask whether social risk is always aversive, answering in the negative and identifying factors that can eliminate, or even change the sign of, the social risk premium. Motivated by the stereotype content model from the social psychology literature, which we argue has straightforward predictions for situations involving social risk (Fiske, Cuddy, and Glick 2007), we focus on two factors: "warmth," synonymous with intent, and "competence." We investigate these factors using a between-subjects experimental design that implements slight modifications of the binary trust game of Bohnet and Zeckhauser across treatments. Our results indicate that having risk generated by another human being does not, on its own, lead to a social risk premium. Instead, we find that a positive risk premium is demanded when a counter-party has interests conflicting with one's own (low warmth) and, additionally, is competent. We find a *negative* social risk premium – i.e., social risk seeking – when the counter-party has contrary interests but lacks competence.

JEL Classification: Z1, C91, D81

Keywords: Social Risk, Social Perception, Intention, Betrayal Aversion, Trust

*a: Einaudi Institute for Economics and Finance, b: Department of Decision Sciences and IGIER, Bocconi University. Miller gratefully acknowledges support from a 2011 grant awarded by the Einaudi Institute for Economics and Finance. Both authors would like to thank seminar participants at EIEF and Bocconi University, as well as conference participants at MBEES in Maastricht (2014), FUR in Rotterdam (2014), and the ESA World Meetings in Honolulu (2014).

1 Introduction

The canonical framework for describing the domain of decision making under risk is that of a lottery, i.e., a probability distribution over consequences. However, a growing body of research investigating how risk and uncertainty affect behavior argues that an individual's willingness to accept risk depends on factors other than merely probabilities and consequences. For example, several studies have noted that the source of risk, e.g., whether or not risk exposure is voluntary or not, affects decision-making and, at the same time, does not fit neatly within the consequentialist lottery framework (Loewenstein, Weber, Hsee, and Welch 2001; Slovic 1987).

In this paper, we focus on one source of risk that has recently captured economists' attention: "social risk." A decision maker faces social risk when another human being is the primary source of uncertainty (Bohnet et al. 2008). In a seminal contribution, Bohnet and Zeckhauser (2004) use laboratory experiments controlling for many plausible extraneous factors—central among them, distributional preferences and ambiguity—to demonstrate that people treat social risk differently than other, inanimate, sources of risk, such as an (ambiguous) randomizing device.[1] Specifically, in a situation involving trust, Bohnet and Zeckhauser find that individuals demand a premium in the probability of receiving a favorable outcome in order to have uncertainty resolved by a human agent rather than by a randomizing device.[2] This *social risk premium* has been attributed to *betrayal aversion* because it can be explained by individuals anticipating an additional disutility when an unfavorable outcome is chosen by a human agent, who can betray someone's trust, rather than a randomizing device, which cannot (Bohnet and Zeckhauser 2004). The betrayal aversion phenomenon has subsequently been documented in several separate studies involving trust that have been conducted across a variety of cultures (Aimone and Houser 2012; Bohnet et al. 2008; Bohnet, Herrmann, and Zeckhauser 2010; Fetchenhauer and Dunning 2009, 2012).

An early conjecture made, but not directly tested, by Bohnet and Zeckhauser themselves (2004, p. 478) is that the social risk premium may simply be one manifestation of a more general aversion to social sources of risk that is driven by a basic desire to avoid relinquishing control to another human agent. More recently, this conjecture has found apparent empirical support in a related setting (Bartling, Fehr, and Herz 2013). The importance of this conjecture stems from the fact that if the social risk premium is driven only by an intrinsic preference for control, then the scope for the study of social risk to contribute to our understanding of behavior would be limited. In this

[1] At the time participants make their choices they are given no specific information about the randomizing device used. Consequently, one may interpret the device as a source of Knightian uncertainty (Knight 1921), i.e., ambiguity.

[2] Bohnet and Zeckhauser (2004) elicit probabilities from first movers in a binarized version of the trust game experimental paradigm of Berg, Dickhaut, and McCabe (1995); the elicitation amounts to a quasi-strategy method applied to pooled second-mover responses. For earlier versions of incentivized experiments involving trust see Camerer and Weigelt (1988) and Fehr, Kirchsteiger, and Riedl (1993).

case, the impact of social risk would be quantitative and not qualitative, i.e., the presence of social risk would have the effect of merely biasing upward any estimate risk aversion, and therefore any further study of its role in decision making would only improve the precision of estimates, rather than uncover factors that may moderate, or even reverse, its effect on risk attitudes.

In this study we design an experiment to test whether the impact of social risk on behavior is indistinguishable from a simple intrinsic preference for control. We find clear evidence to the contrary: the prospect of relinquishing control to another human agent does not necessarily lead people to demand a premium for social risk. In particular, we find that for situations in which a human agent who acts with the intention to betray *cannot do so effectively*, individuals deciding whether or not to expose themselves to this social source of risk behave as if they *prefer* it to a risk stemming from an inanimate source. Accordingly, a deeper understanding of how the presence of social risk interacts with contextual factors such as intention and the opportunity to betray may ultimately yield valuable insights into how behavior varies across consequentially identical situations.

Our approach is motivated by a theoretical framework from the social psychology literature that, in our view, can explain both the betrayal aversion phenomenon and predict new patterns. A recent theory of how people perceive strangers and form stereotypes about social out-groups, the *stereotype content model* (Fiske, Cuddy, Glick, and Xu 2002), incorporates insights from earlier work on social perception (Asch 1946; Bales 1950; Rosenberg, Nelson, and Vivekananthan 1968) and organizes patterns of how stereotypes operate in different cultures (Cuddy et al. 2009; Fiske et al. 2002). The basic finding is that the personality impressions that people form about other individuals, how they construe behavior, and the stereotypes they hold about members of other groups, can be categorized into two factor dimensions, which have been labeled *warmth* and *competence* respectively (Fiske et al. 2007). Here, warmth is largely synonymous with intent—positive (negative) intentions being identified with high (low) warmth.[3] The impression of another agent's personal warmth serves as a cue to what the other agent's possible goals are with respect to the self, while the impression of the agent's competence is thought to serve as a cue to the agent's ability to carry out those goals.[4]

Context can determine the formation of these personality impressions, which can, in turn, predict affective reactions. Social out-groups who compete for resources and successfully control them in their own favor tend to be viewed has having low warmth and high competence, which, in

[3] In their paper introducing the stereotype content model, Fiske et al. (2002) present a model of people as having pragmatic/consequentialisist objectives when dealing with strangers: "when people meet others as individuals or group members, they want to know what the other's goals will be vis à vis the self or in-group and how effectively the other will pursue those goals. That is, perceivers want to know the other's intent (positive or negative) and capability; these characteristics correspond to perceptions of warmth and competence respectively."

[4] In the fields of management and sociology there is a similar two-factor definition of trust: trusting another human agent involves (1) trusting in an agent's competence, and/or, (2) trusting in an agent's intentions (Nooteboom 2002).

turn leads them to be envied or perceived as a threat. The prospect of ceding control to potentially competitive agents, such as in a trust game, may therefore generate a negative affective reaction (Fiske et al. 2002). On the other hand, if these same agents experience personal misfortune, a positive affective reaction may result if it is perceived as a removal of a social threat; this reaction has been associated with an emotion known as *schadenfreude* (Cikara and Fiske 2013). Since the probability of such misfortune increases in incompetence when an agent has control, the knowledge that the agent, competing for the same resources, may not be able to effectively pursue his own interests could counterbalance or even overcome the negative affect associated with the prospect of ceding control to him.[5]

Assessing the risk generated by interacting with another human agent in an experimental social dilemma, such as the trust game, also involves an act of social perception in order to anticipate the behavior of one's counterpart. While an anonymous laboratory setting appears to provide little scope for forming personality impressions or using stereotypes, evidence suggests that the perceptions and stereotypes people form about others can be driven solely by context, in particular the degree of competition and the relative control over resources one's counterpart has vis-à-vis the self (Cikara and Fiske 2013; Fiske et al. 2002). We propose that individuals involved in social dilemmas respond to changes in these contextual details in a pattern consistent with the stereotype content model. An individual may exhibit an aversion to the possibility of betrayal and demand a premium to expose themselves to social risk because of the negative *affect* associated with facing the potentially threatening intentions of a human agent who competes for resources and can competently control them in his favor. By contrast, if a human agent can control resources but cannot competently do so in his own favor, then the individual is protected from the influence of any threatening intentions the agent may have. This may, in turn, generate a positive affect towards the prospect of being exposed to social risk, and partially offset or even change the sign of the social risk premium. In the current study, we test for this predicted negative relationship between competence and the social risk premium in a two-player setting, holding constant the competition for resources embodied by monetary incentives.

Toward this end, we implement a between-subjects experiment comprised of four separate treatments: the Random Device (RD) treatment, which is a version of the Risky Dictator game of Bohnet and Zeckhauser (2004); the Human (H) treatment, which is the Binary Trust Game of Bohnet and Zeckhauser (2004); as well as two additional treatments intended to test our specific

[5] In Section 2 we model this by assuming that when a competitive agent is seen to be incompetent, this leads decision makers to find the social risk generated by the agent as less aversive, i.e., decision makers experience an increase in their decision utility for exposing themselves to social risk. Alternatively, one can model the relative relief decision makers face when they anticipate their affective reaction associated with the possible misfortune of the agent (*schadenfreude*) as an increase in utility in that contingency, without any influence on utility in other contingencies.

hypotheses. These latter two treatments are the Human-Unforeseeable (H-UF) treatment and the Human-Unaware (H-UA) treatment. In each experimental session, each participant is randomly assigned to exactly one of these four treatments and is not informed of the existence of the other three treatments. Each treatment features a suitably modified version of the conditional Binary Trust Game of Bohnet and Zeckhauser (2004), which is one-shot, anonymous, and uses payoff parameters identical to the original study.

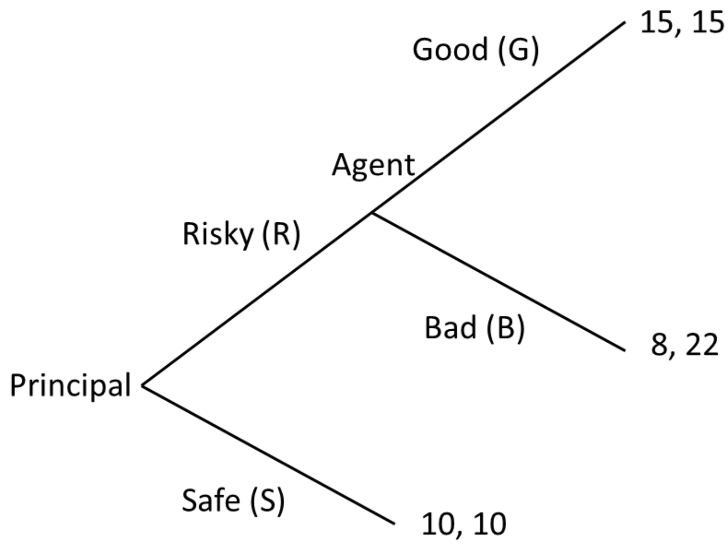

Figure 1: *The Game*

The experimental treatments are all based on the game presented in Figure 1. Each treatment, involves a principal (first mover) and an agent (second mover) being randomly matched. The principal decides between R (risky) and S (safe). If she selects S she receives 10 Euros and the agent receives the same, regardless of his choice. If the principal selects R, then if the agent chooses G (or has G selected for him) each player receives 15 Euros, while if the agent chooses B (or has B selected for him), the principal receives 8 Euros, while the agent receives 22 Euros.[6]

We use the strategy method for the agents, who choose between G (good) and B (bad) before knowing if the principal has selected R (risky) or S (safe). The principal's decision between R and S is made conditionally. We ask the principal to state the minimum acceptable probability (MAP), p, of their agent choosing G that would make the principal prefer R to S. If the relative frequency of agents choosing G (or having G selected for them) is greater than or equal to p, the principal commits to choosing R, otherwise the principal chooses S. This mechanism provides the principal proper incentives to truthfully report his or her MAP under mild assumptions which we discuss

[6] In the actual experiment we used neutral wording to label players and actions. See Section A.3 of the appendix for details.

later.

In treatment RD, the agent is entirely passive. A randomizing device with a fixed probability selects between G and B on behalf of each agent before the principal chooses. Treatment H is identical, except the agent makes decisions for himself. Treatments RD and H together essentially replicate the canonical betrayal aversion setup of Bohnet and Zeckhauser (2004), although Treatment RD is slightly modified to keep all treatments parallel, i.e., the MAP is compared to the relative frequency of G in the realizations.[7]

Treatment H-UA is designed to test if the mere fact that another human controls the outcome leads individuals to demand a premium for exposing themselves to social risk. Treatment H-UA is identical to treatment H except players do not have common knowledge of the game. If the principal selects R, the agent chooses between G and B while being unaware that he is in a game with earnings consequences arising from his choice. Because the principal is informed of this, while the agent is controlling the outcome, from the principal's perspective the agent is not a competitor with conflicting interests. This suggests that the principal may treat the agent as equivalent to a randomizing device.

Our final treatment, H-UF, is again identical to treatment H except the actions available to the agent are relabeled with new symbols so that he is unsure which of the alternative symbols implement which actions, and thus he cannot foresee the consequences of his own choices. Thus, the agent knows all of the outcomes of the game, can have preferences and intentions over these outcomes, and determines the outcome with his choice. However, the agent's competence is low as he is effectively prevented from reliably implementing his preferred outcome. Comparing MAPs in the H-UF treatment to the MAPs in H will provide evidence on how counter-party competence affects the the principal's attitude toward social-risk.

As a preview of our results, we find that the relationship between social-risk attitudes and human agency is more subtle than previously thought. First of all, we essentially replicate Bohnet and Zeckhauser's original findings: MAPs in the H treatment are higher than those in the RD treatment, suggesting individuals are willing to pay a premium to have outcomes decided by chance rather than a human co-player. On the other hand, by comparing MAPs in the H-UF and RD treatments, we find the opposite effect of social risk. In a finding suggestive of *schadenfreude* (Cikara and Fiske 2011, 2013), MAPs are *lower* in H-UF than in RD, which indicates that participants are willing to accept a probability *discount* to have outcomes determined by another agent who cannot effectively pursue his own self-interest (incompetent), instead of an inanimate randomizing device. Finally,

[7] In the Risky Dictator Game of Bohnet and Zeckhauser (2004) participants' MAPs were compared to the objective probability of a randomizing device while in the Binary Trust Game MAPs were compared to a realized relative frequency. The results here replicate the Betrayal aversion phenomenon in this modified setting.

comparing MAPs between H-UA and RD we find no difference. This last pattern, along with the result that MAPs are significantly lower in H-UF than in RD, both suggest that an aversion to social risk is not driven solely by an aversion to having another human agent influence outcomes. Further, because RD and H-UA differ primarily in whether agents know their monetary interests are in conflict with the principal's (low warmth), this non-result suggests that agents' knowledge of their conflicting interests does not contribute substantially to social-risk aversion exhibited by principals.

The remainder of the paper proceeds as follows. In Section 2, we describe the background to our hypotheses in detail. In Section 3, we provide details on our experimental design. In the subsequent section we present the results of our experiments. Section 5 concludes.

2 Hypotheses

Consider an individual, whom we term the principal, choosing between S and R in Figure 1. Suppose she attaches utilities to the three outcomes, denoted as U_S, U_G and U_B. If outcomes are determined by a randomizing device as in Treatment RD and p is the probability of the good outcome, the ex-ante (expected) utility of choosing R is given by, $p \cdot U_G + (1-p) \cdot U_B$. The minimum acceptable probability (MAP) is the p which makes the principal indifferent between R and S. For Treatment RD, this MAP is given by:

$$MAP_{RD} = \frac{U_S - U_B}{U_G - U_B}$$

When the choice of R involves relinquishing control to another human agent, as in Treatments H, H-UA, and H-UF, this may create an additional non-monetary cost (or benefit). The ex-ante utility will instead given by $p \cdot U_G + (1-p) \cdot U_B - c$, with $c \in \mathbb{R}$. Therefore the MAP will depend on c, and will satisfy:

$$MAP(c) = MAP_{RD} + c/(U_G - U_B)$$

Because the principal has no information about the agent she is matched with, we propose that she will use contextual cues in Treatments H, H-UA, and H-UF to help her predict both the goals of the agent and the agent's ability to achieve these goals. How the principal uses these cues will follow the patterns outlined in the stereotype content model.

In Treatment H, the human agent is competing for resources and can effectively control them. In analogous social situations, human counterparts are typically perceived as posing a threat, engendering a negative affective reaction and/or stereotype (Cuddy et al. 2009; Fiske et al. 2002). In an anonymous laboratory situation a similar affective response may be evoked. This negative affect may make choosing R feel additionally aversive relative to a randomizing device, with the principal experiencing a non-monetary cost for choosing R, $c > 0$. This implies $MAP_{RD} < MAP_H$, which is consistent with the behavior observed in studies where there is a threat of betrayal.

In Treatment H-UF, the human agent is competing for resources but he cannot effectively control them, i.e., the principal is protected from any threatening intentions the agent may have. In analogous social situations, when a competitor suffers misfortune because he is not completely in control of outcomes, this is typically associated with a reduction in the perceived threat level and may even evoke positive affect (e.g., *schadenfreude*, Cikara and Fiske (2011, 2013)). In an anonymous laboratory situation, hampering a competitor's ability to effectively control outcomes may conjure a similar positive affective response. Relative to a random device, then, choosing R may feel less aversive so that the principal experiences a non-monetary benefit for choosing R, $c < 0$. This implies $MAP_{H\text{-}UF} > MAP_{RD}$, which, if true in the data, would: (1) provide evidence that the premium demanded for social risk in the trust game is unrelated to a general reluctance to cede control to another human agent; and (2) suggest that when people feel "protected," they are more willing to expose themselves to social risk than the objective level of risk justifies.

In Treatment H-UA, the human agent is unaware that his actions influence his own payoffs, let alone the payoffs of the principal. Without the typical features of competition and control, there is no reason for the principal to assess the goals of the agent or the agent's ability to achieve these goals, and therefore the stereotype content model has no clear prediction. The agent is acting essentially as a human randomizing device, and therefore if the mere fact that an agent has control over outcomes is not a driver of social risk, we should expect $MAP_{H\text{-}UA} = MAP_{RD}$.

3 Design

All experimental sessions were conducted at Bocconi University. The participants consisted of 158 undergraduate students recruited from the Bocconi Experimental Laboratory for the Social Sciences (BELSS) on-line subject recruitment system and 11 graduate students who were recruited individually via email.[8] Neutral wording for roles and outcomes were used in the experimental instructions (see Appendix B). Here, we use more descriptive terminology for ease of exposition.

All undergraduate students were assigned the role of principal (described below). Each undergraduate student participated in a single session of approximately 27 students in "Room X" (6 sessions total) and was randomized into exactly one of four treatments, again described below. Each graduate student in "Room Y was assigned the role of agent and was assigned to all four treatments of each "Room X" session.[9] Our focus will be on participants assigned the role of

[8]The website and database is administered by Sona-Systems. The average age of those in our sample was 21 students.

[9]The agents submitted decisions in each treatment in the following order: H-UA, H-UF, H, and RD. Agents

principal, all of whom were informed only that they were matched with participants in Room Y, on the fifth floor of the building, whose decisions could potentially affect their own outcomes.

Upon arrival in Room X, principals selected one card from a shuffled deck of cards. The chosen card revealed a "code" number, from 1 to 27. Participants were instructed to sit in the private carrel of the laboratory corresponding to their code numbers. Once all participants were seated they were immediately informed that the amount they would be paid at the end of the session would be between 8 Euros and 15 Euros. Furthermore, they were informed that how much they would be paid depended on a single choice that they would make together with a choice of the agent in Room Y with whom they had already been matched, and that they should therefore listen and read carefully.[10]

Next the experimenter read a welcome script (See Appendix B.2). Principals were informed that they were in Room X and the code number they drew upon entering the room determined the agent they were matched with in Room Y. They were told that all participants' identities would be kept anonymous.[11] Next, principals were given an overview of the experimental session and told that: (1) they would receive detailed instructions for them to read to themselves; (2) they would answer a short quiz intended to check their understanding of the instructions; (3) we would check their responses to the quiz; (4) they would respond to the single "Key Question" (which was our outcome measure of interest); (5) while we were matching their responses to that of the other player in the role of agent from Room Y, they would fill out a survey; and finally, (6) that they would be paid.

For each X-session block, the shuffled "code" numbers implemented a permuted block randomization with a uniform allocation of participants into treatments.[12] Principals were not aware they were assigned to a treatment nor that there were other treatments.[13]

The Treatments Each treatment had a common underlying structure in terms of how payoffs depended on the choices of the principal and the agent.[14] In all treatments, the principal made a decision which determined whether S, the safe option, or R, the risky option, was selected. If the principal "chose" option S, this yielded a certain outcome for the pair, and the experimental

received payments from all four treatments, but principals were not informed of this.

[10]The participants in Room Y decided in advance using the strategy method as is detailed in the treatment description below.

[11]It was important to match before they choose rather than after, to make clear the decision is coming from their assigned agent rather than the outcome being a draw from a pool of already determined decisions.

[12]For sessions of 27 participants the allocations in each treatment were 6,7,7,7. For sessions of 26 participants the allocations in each treatment were 6,6,7,7.

[13]The agents were provided identical instructions as the principals, and given a separate sheet to submit their responses. In Treatment Unaware the agent had no instructions and only a choice

[14]Payoffs were denominated in euros, the numerical amounts were identical to the original experiment Bohnet and Zeckhauser (2004) as well as the as well as the cross-cultural studies Bohnet et al. (2008) and Bohnet et al. (2010).

earnings were 10 euros to the respective players. If the principal "chose" option R, then the payoffs depended on the choice of the agent. If the agent chose option G, then experimental earnings were 15 euros for both players. If the agent chose option B, then the agent received 22 euros in experimental earnings and the principal received 8 euros in experimental earning.[15] In all treatments the agent's choice between options G and option B conditional on the principal choosing alternative R was determined before the principal chose (strategy method). In each treatment, the principal's choice between alternatives S and R was implemented as in the trust game of Bohnet and Zeckhauser (2004) with instructions that closely followed those reported in Bohnet et al. (2008). For each subject in the role of principal, we elicited the value p which was described as their *minimum acceptable probability* (MAP) of being matched with an agent choosing option G that would lead them to choose alternative R over alternative S. If the percentage of subjects in Room Y choosing G, p^*, was greater than or equal to p, then the principal committed to choosing alternative R and payoffs would be determined by the choice of the agent with whom the principal had (already) been matched.[16] If we assume that the difference in utility between S and R as a function of p^*, $\Delta U(p^*) = U(R, p^*) - U(S, p^*)$, is strictly increasing in p^* on $[0, 1]$ and satisfies $\Delta U(1) > 0 > \Delta U(0)$, then for the principal reporting $p = MAP$, where $MAP := \inf\{p^* > 0 : \Delta U(p^*) > 0\}$, is a unique weakly dominant strategy.[17]

The random matching made each treatment structurally identical for a principal who cares only about the probability that the agent "chooses" G. The distinguishing feature between treatments was the mode in which the agent "chooses" between options G and B. These differences are presented below for each treatment:

Human (H): Each agent decided directly between option G and option B. This is the Trust Game from Bohnet and Zeckhauser (2004).[18]

Human-Unaware (H-UA): Each agent was presented with a row of 17 cells and asked to choose one cell. Agents were not aware that they were playing a game with another player, and therefore they did not know that in each cell there was either a G or a B, and that the cell they selected would determine the payoff for themselves and for their respective principals.

[15] In the experiment, we use the more neutral letters J and K for options G and B, respectively, primarily because they are not present in the Italian alphabet and thus participants were unlikely to have associations with these letters.

[16] By stating the question in this way, our design, and that of Bohnet and Zeckhauser (2004), implicitly assume that principals would like to statistically discriminate against agents. If principals did not wish to discriminate they could simply report $p = 0$ or $p = 1$, and 4 out of 158 subjects did this.

[17] Alternatively, using the justification present in Bohnet et al. (2008), if we assume players believe that p^* is drawn from a distribution where the support contains a neighborhood of their MAP, then reporting p equal to their MAP is strictly dominant as this is equivalent to a Becker-DeGroot-Marshak (BDM) elicitation procedure with p^* generated by the collective behavior of agents in Room Y (Becker, Degroot, and Marschak 1964).

[18] The instructions for Treatment H can be found in Appendix Section B.3.

The principals were aware of this.[19]

Human-Unforeseeable (H-UF): Each agent was presented with a row of 17 cells and asked to choose one cell. Agents were aware that in each cell there was either a G or a B, and that the cell they selected would determine the payoff for themselves and for their principals. The agent could not see the contents of each cell and thus could not foresee whether G or B would be selected. Principals also could not see the contents of the cells. Both principals and agents were aware of this.

Random Device (RD): Each agent was presented with a row of 17 cells and a randomizing device determined one of the cells at random.[20] Agents were aware that in each cell there was either a G or a B, and that the cell the randomizing device selected would determine the payoff for themselves and for their principals. The agent could see the contents of the cells, but the principal could not. Both principals and agents were aware of this.[21]

4 Results

In Table 1 it is evident that the average MAP in our H treatment is similar to those reported for other western countries (Switzerland and the United States) in the most directly related treatment of Bohnet et al. (2010).

Our RD treatment employs a randomizing device, rather than the Room Y co-player, to determine the selection between G and B. This treatment is analogous to the Risky Dictator Game of Bohnet et al. (2010).[22] In Table 2 we see that the MAPs in our study are not as low as in Bohnet et al. (2010). A potential explanation for this is that the risk generated from being randomly matched to an outcome from a sample of realizations of a random device is perceived differently than the equivalent risk of receiving a single outcome directly from the device itself.[23] The results

[19] The instructions for the principals were identical to the instructions in treatment H, except for the existence of the 17 cells. The choice of 17 cells was made so that a uniform distribution over G and B was unlikely to be focal from the perspective of the principals. The decision not to include the 17 cells in treatment H when implementing the design was made to avoid the risk that subjects could become confused with the introduction of a device that transparently serves no purpose. While the visual representation of the instructions is slightly different, there is some indication that it is justifiable to assume that these differences are slight, and do not confound our results: the average MAP in treatment RD is only slightly higher than that in the study of Bohnet et al. (2010), with a similar population, though the comparison is not based on random assignment (see Table 2).

[20] The selection was made using `www.random.org`.

[21] Unlike the Risky Dictator game reported in BZ, which treated p^* as an ex-ante probability of a random device, in all treatments of our study p^* was equal to the empirical relative frequency of G choices, even if the G choices were determined by realizations of the randomizing device.

[22] In our study, however, MAPs are elicited in a more parallel fashion across treatments: MAPs are compared to the actual empirical relative frequency of G choices (p^*) made by a randomizing device on behalf of the agents, rather than to the ex-ante probability of G being selected by the randomizing device itself.

[23] In Bohnet et al. (2010) the device was a an urn with an unknown distribution of balls, in our study it is the third party uniform random number generator based on atmospheric noise (`www.random.org`) and an unknown distribution

Table 1: *Minimum Acceptable Probabilities in Treatment H (Mean,* Median, *[N])*

	ALL	Men	Women
Milan	0.54	0.51	0.59
	0.55	*0.45*	*0.60*
	[38]	[25]	[13]
Switzerland	0.51	0.46	0.62
	0.55	*0.48*	*0.60*
	[25]	[18]	[7]
United States	0.54	0.50	0.61
	0.50	*0.50*	*0.72*
	[31]	[19]	[12]

Data from Switzerland and the United States are from the Trust Game of Bohnet et al. (2010).

in Table 2 suggest that the Betrayal Aversion effect itself could be (partially) driven by the effect of these differences.[24]

Table 2: *Minimum Acceptable Probabilities in Treatment RD (Mean,* Median, *[N])*

	ALL	Men	Women
Milan	0.46	0.43	0.54
	0.50	*0.40*	*0.60*
	[40]	[27]	[13]
Switzerland	0.40	0.33	0.48
	0.42	*0.30*	*0.50*
	[24]	[13]	[11]
United States	0.32	0.28	0.38
	0.29	*0.29*	*0.35*
	[29]	[16]	[13]

Data from Switzerland and the United States are from the "Risky Dictator Game" of Bohnet et al. (2010).

Comparisons Across Treatments Having established broad comparability with previous results we now turn to comparing patterns across treatments within our own study. The main test we employ is the permutation test, as the random assignment was conducted as a permutated block design with session strata. In Table 3 we present simple means of participants' MAPs for each our treatments, separately. Comparing the average MAP in our H treatment to MAPs in our RD treatment, we find evidence consistent with previous results on betrayal aversion: participants are

of Gs and Bs to select from. Some potential sources of the perceptual difference could be: (1) whether chance is realized ex-ante to the decision or ex-post (citations needed); (2) the mechanism itself (matching vs. draws), which could be a real issue because that means that the measurement device creates an artifact (our evidence suggests it is not an artifact, which is a potentially important ancillary result); (3) vulnerability to experimenter manipulation.

[24] Or if one is unwilling to view our 1-17 representation as innocuous, then the difference could be there.

willing to pay a 7 percentage point MAP premium ($p = 0.048$, one-tailed permutation test) to have outcomes determined by random chance instead of a human co-player who has conflicting monetary interests and can perfectly implement his desired outcome. This result provides additional evidence for the robustness of the original BZ findings and, at the same time, provides reassurance that our experimental design and subject pool are reasonable.

Table 3: *Minimum Acceptable Probabilities Across Treatments (Mean,* StdDev*, [N])*

	H	H-UF	H-UA	RD
MAP	0.54	0.37	0.50	0.47
	0.03	*0.04*	*0.04*	*0.03*
	[38]	[40]	[38]	[40]

Our main finding stems from comparing MAPs in RD, where the decision-maker is an inanimate randomizing device, to MAPs in our H-UF treatment where the decision-maker is a human agent with conflicting monetary interests but is incompetent, i.e., cannot effectively implement his own self-interest. Here, we find that participants are willing to pay a substantial ten percentage point premium ($p = 0.032$, one-tailed permutation test) to let the *human co-player* decide the outcome, rather than the randomizing device. Additionally, to the extent that treatment H-UA can be treated as similar to treatment RD (addressed below), we also find that participants are willing to pay an even more substantial 13 percentage point premium ($p = 0.004$, one-tailed permutation test) to let the incompetent *human co-player* decide the outcome, rather than the human randomizing device. One explanation for this finding is anticipated *schadenfreude*: players in the role of principal gain utility from the possibility that their hamstrung opponent who may intend to harm them, could actually harm themselves accidentally. Anecdotally, this is a familiar plot device in many comedies. A somewhat opposite story may also be at work, of course: a given percentage of bad outcomes may actually be due to individuals who intend to do good and only mistakenly do harm. Our experiment cannot precisely separate these two stories as they have observationally equivalent predictions with respect to MAPs in our H-UF treatment. However, they both have an interesting implication: a bit of expected incompetence in a population can increase trusting behavior.

Finally, comparing MAPs in our RD treatment to those in our H-UA treatment provides evidence that there is nothing special about the uncertainty stemming from human behavior *per se*. The decision-maker in H-UA is human while the decision-maker in RD is inanimate, yet there is no difference in the MAPs principals demand, on average. Table 4 presents significance tests for all pairwise comparisons among all of our treatments.

Turning from simple means to distributions, in Figure 2 we present histograms, overlaid with

Table 4: *Pair-wise between-treatment differences in the mean MAPs and p-values for permutation tests (one-tailed). The permutations test is stratified at the session level to match the permuted block design (See Appendix Section A for details).*

Comparison	Difference	P-Value
H vs. RD	0.08^{**}	.0483
H vs. H-UF	0.17^{***}	.0002
H vs. H-UA	0.04	.1883
RD vs. H-UA	−0.04	.7881
RD vs. H-UF	0.09^{**}	.0325
H-UA vs. H-UF	0.13^{***}	.0038

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided, right)

kernel density estimates, of participants' MAPs for each of our treatments separately. An important point to notice is that in all treatments a wide range of MAPs are reported. This is important because it makes it unlikely that our results are driven by a few outliers. A second point to notice is that the histograms and kernel density estimates tend to corroborate the story gleaned from comparing means. For example, the distribution of MAPs in our H-UF treatment are essentially a leftward shift of the MAPs in our H treatment. Low MAPs (more trusting behavior) are more prevalent when we introduce noise into the mapping between co-player's action and outcomes, bringing down the average MAP for the H-UF treatment.

5 Concluding Remarks

In this study we experimentally investigated how two factors suggested by the social psychological literature on person perception (Asch 1946; Fiske et al. 2007, 2002; Rosenberg et al. 1968), warmth and competence, affect individuals' attitudes toward social risk. To facilitate comparison with previous literature, our study features the binary trust game common in several recent studies of betrayal aversion, starting with Bohnet and Zeckhauser (2004). We constructed a simple model incorporating warmth and competence into our consideration of social-risk attitudes and used it to generate testable hypotheses about how these two factors might affect attitudes toward social risk.

Our results suggest that social risk engenders more nuanced attitudes than previously thought. We start out by replicating the by-now standard *betrayal aversion* finding that individuals may be willing to pay a premium to have outcomes decided by chance rather than a human co-player. We go on to show that, consistent with our predictions, this need not always be the case. We compare MAPs in a situation where an agent lacks warmth but also lacks competence (H-UF) to a situation where outcomes are decided by pure chance (RD and H-UA) and find an opposite attitude toward social risk. In a finding reminiscent of *schadenfreude* (Cikara and Fiske 2011, 2013), participants are

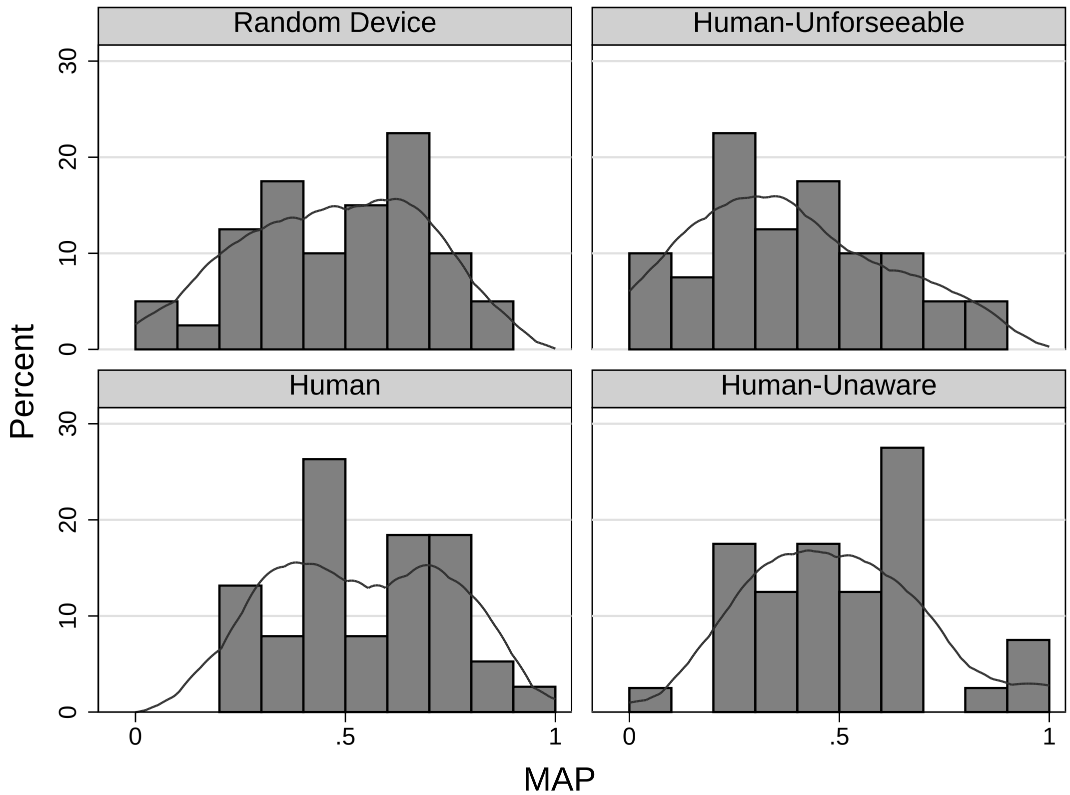

Figure 2: *Histogram of MAPs by treatment (with Kernal Density)*

willing to pay a premium to have outcomes determined by a hamstrung or incompetent opponent rather than by chance. Finally, comparing MAPs between our H-UA and RD treatments we find evidence consistent with there being little or no pure human co-player effect with respect to the social risk premium. Principals do not report significantly higher MAPs in H-UA than in RD.

Because not only the magnitude, but also the sign, of the social risk premium may change across situations with identical outcomes, the presence of social risk may produce qualitatively different behavior across otherwise-identical choices. For example, when the resolution of uncertainty depends on another (competitive) human's actions, individuals may appear risk-seeking in one environment while appearing risk-averse in an apparently identical environment, depending on the competence of the humans on whom uncertainty depends.[25] Moreover, eliminating social risk by resolving uncertainty with the use of an inanimate randomizing device may either enhance or undermine trusting behavior, depending again on the competence of the human counter-parties requiring trust.

[25]In the present experimental study, incompetence was externally manipulated, and was not remediable. It is unclear what may happen when incompetence is endogenous.

A potentially interesting implication of our finding that an agent's incompetence appears to make people more tolerant of social risk is that, knowing this, a strategic agent may feign incompetence to elicit trust. Such "strategic incompetence" requires that agents correctly anticipate the effect of incompetence on principals' attitudes toward social risk that we document so that it is not clear how large a role feigned incompetence may play in actual behavior. We leave the study of strategic incompetence for future research.

References

AIMONE, J. A. AND D. HOUSER (2012): "What you don't know won't hurt you: a laboratory analysis of betrayal aversion," *Experimental Economics*, 15, 571–588.

ASCH, S. E. (1946): "Forming impressions of personality," *Journal of Abnormal and Social Psychology*, 42, 248–290.

BALES, R. (1950): "A set of categories for the analysis of small group interaction," *American Sociological Review*, 15, 257–263.

BARTLING, B., E. FEHR, AND H. HERZ (2013): "The Intrinsic Value of Decision Rights," University of Zurich, Department of Economics Working Paper No. 120. Available at SSRN: http://ssrn.com/abstract=2255992.

BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1964): "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226–232.

BERG, J., J. DICKHAUT, AND K. MCCABE (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122–142.

BOHNET, I., F. GREIG, B. HERRMANN, AND R. ZECKHAUSER (2008): "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States," *American Economic Review*, 98, 294–310.

BOHNET, I., B. HERRMANN, AND R. ZECKHAUSER (2010): "Trust and the Reference Points for Trustworthiness in Gulf and Western Countries," *Quarterly Journal of Economics*, 125, 811–828.

BOHNET, I. AND R. ZECKHAUSER (2004): "Trust, risk and betrayal," *Journal of Economic Behavior & Organization*, 55, 467–484.

CAMERER, C. AND K. WEIGELT (1988): "Experimental Tests of a Sequential Equilibrium Reputation Model," *Econometrica*, 56, 1–36.

CIKARA, M. AND S. T. FISKE (2011): "Stereotypes and Schadenfreude: Affective and Physiological Markers of Pleasure at Outgroup Misfortunes," *Social Psychological and Personality Science*.

——— (2013): "Their pain, our pleasure: stereotype content and schadenfreude," *Annals of the New York Academy of Sciences*, 1299, 52–59.

CUDDY, A. J. C., S. T. FISKE, V. S. Y. KWAN, P. GLICK, S. DEMOULIN, J.-P. LEYENS, M. H. BOND, J.-C. CROIZET, N. ELLEMERS, E. SLEEBOS, T. T. HTUN, H.-J. KIM, G. MAIO, J. PERRY, K. PETKOVA, V. TODOROV, R. RODRIGUEZ-BAILON, E. MORALES, M. MOYA, M. PALACIOS, V. SMITH, R. PEREZ, J. VALA, AND R. ZIEGLER (2009): "Stereotype content model across cultures: Towards universal similarities and some differences," *British Journal of Social Psychology*, 48, 1–33(33).

FEHR, E., G. KIRCHSTEIGER, AND A. RIEDL (1993): "Does Fairness Prevent Market Clearing," *Quarterly Journal of Economics*, 108, 437–459.

FETCHENHAUER, D. AND D. DUNNING (2009): "Do people trust too much or too little?" *Journal of Economic Psychology*, 30, 263–276.

——— (2012): "Betrayal aversion versus principled trustfulness: How to explain risk avoidance and risky choices in trust games," *Journal of Economic Behavior & Organization*, 81, 534–541.

Fiske, S. T., A. J. Cuddy, and P. Glick (2007): "Universal dimensions of social cognition: warmth and competence," *Trends in Cognitive Sciences*, 11, 77–83.

Fiske, S. T., A. J. Cuddy, P. Glick, and J. Xu (2002): "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition," *Journal of Personality and Social Psychology*, 82, 878–902.

Knight, F. (1921): *Risk, Uncertainty, and Profi*, Houghton Mifflin Company.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch (2001): "Risk as feelings," *Psychological Bulletin*, 127, 267–286.

Nooteboom, B. (2002): *Trust: Forms, foundations, functions, failures and figures*, Cheltenham, U.K.,& Northampton, MA: Edward Elgar Publishing.

Rosenberg, S., C. Nelson, and P. S. Vivekananthan (1968): "A multidimensional approach to the structure of personality impressions," *Journal of Personality and Social Psychology*, 9, 283–294.

Slovic, P. (1987): "Perception of risk," *Science*, 236, 280–285.

A Appendix: Supplementary Tables & Figures

In Table 5 we report the output of the permutation test for the difference in means, for each between-treatment comparison. The permutation follows the permuted block design of the experiment, where permutations are stratified at the experimental session level. The second column lists the estimated difference between treatments. The third column counts the number of permutations (out of 100,000) where the difference was at least a large as the estimated difference. The fourth column lists the approximate p-values, the proportion of the permuted data where the difference is at least as large as the estimated difference.[26] The 95 percent confidence interval pertains to the p-value, it is a binomial (Clopper-Pearson) confidence interval based on the 100,000 realizations from the permutation scheme.

Table 5: *For each pair-wise between-treatment difference in means, the results of the permutation test are reported below.*

Comparison	Difference	Count	P-Value	St. Err.	[95% Conf.	Interval]
H vs. RD	0.08^{**}	4833	0.048	0.001	0.047	0.050
H vs. H-UF	0.17^{***}	22	0.000	0.000	0.000	0.000
H vs. H-UA	0.04	18831	0.188	0.001	0.186	0.191
RD vs. H-UA	-0.04	78808	0.788	0.001	0.786	0.791
RD vs. H-UF	0.09^{**}	3253	0.033	0.001	0.031	0.034
H-UA vs. H-UF	0.13^{***}	381	0.004	0.000	0.003	0.004

100,000 Permutations (player strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided, right)

In Table 6, for the purposes of comparison, we report the outcome of the T-test, where the distribution of mean MAPs in each session is assumed to be normal.

Table 6: *Below we can see that under the assumption of normality, the p-value of the T-test yields a close approximation of the exact p-value*

Comparison	Permutation	T-test
H vs. RD	.0483	.0556
H vs. H-UF	.0002	.0005
H vs. H-UA	.1883	.2030
RD vs. H-UA	.7881	.7665
RD vs. H-UF	.0325	.0317
H-UA vs. H-UF	.0038	.0063

[26] There exists exact p-values for this test, and the Monte-Carlo permutations can approximate them to arbitrary precision.

B Appendix: Experimental Procedures & Instructions

B.1 Procedures

Phase 1 One week before the main experiment sessions (phase 2) 10 students in the graduate program at Bocconi University were invited to take part in an experiment in room "Y". In a single session, these students participated as the second mover (Person Y) making a pre-commitment of their choice in response to the choice of the first mover (Person X) in every experimental treatment in the following order (1) Treatment Human-Unaware (H-UA), (2) Treatment Human-Unforeseeable (H-UF), (3) Treatment Random Device (RD), and finally (4) Treatment Human (H)s. For each treatment they were told that they are Person Y but they are to read the instructions for Person X as their instructions describe every aspect of the game.[27] These students returned after phase 2 of the study to receive their payment.

Phase 2 Participants were recruited from the Bocconi University online recruitment website administered by Sona-Systems (http://www.sona-systems.com/). Each session was given a unique title and description to minimize communication between participants.

When participants arrived they waited until all registered students were present and then were invited into the lab all at once. As they walked in, they selected "code" numbers out of a box and were told immediately: "You have been paired with a another participant".[28] Subjects were next instructed to seat themselves in the carrel corresponding to their code number. When they were seated the experimenter began with the "Experimenter Script' presented in Section B.2.

When the script was finished each instruction/decision sheet was folded in half and handed out. In each session there were four sets of instruction/decision sheets, one corresponding to each experimental treatment participants were assigned to. The selection of "code" numbers implemented a permuted block randomization, with a block size of 27 participants and a near-uniform allocation ratio (6,7,7,7).

Participants read the instructions privately and raised their hands to ask clarification questions. When instructions were complete participants filled out a quiz checking their understanding. Next quizzes were collected. Incorrect quizzes were identified by experimental assistants and replaced with a new blank quiz and participants were given an opportunity to ask questions again (the process continued until each participant could demonstrate understanding of the instructions).

After the instruction/decision sheet was collected from each participant they were handed a survey and a receipt form to fill out while the experimenter matched them with Room Y decisions and determined their payment.[29] Next participants were called up one-by-one to be paid based on the choice of the person from Room Y (phase 1) whom they were matched with.

Phase 3 Students from phase 1 (Room Y) returned one week after phase 2 and were paid for each treatment they participated in. For each treatment their earnings from each participant they were matched with were pooled together. They were paid based on a random selection from their pooled earnings from each treatment. Participants in Room X were not informed that the matching and the payment for participants in Room Y would be conducted precisely in this manner.

[27] In Treatment Unaware there were no instructions, subjects were simply asked to choose a box.

[28] This was for the random pairing with Room Y participants.

[29] The specific implementation of the matching was not described to the participants of Room X, only that they were uniquely matched with a student from Room Y. The matching was many-to-one and only participants in Room Y were aware of this.

B.2 Experimenter Script (English Translation): Room X Sessions

1. (Once everyone is seated) Welcome to the study and thank you for participating.

2. First, we ask you to please turn off your mobile devices, not communicate between each other, and leave your desk clear of everything except your student ID and a pen. We will not be using the computers.

3. We will give you a brief overview of the study. It is important that you listen closely. You may ask questions once we have finished reading the instructions (which we will hand to you shortly)

4. This is Room X. When you selected a code number as you walked in this room you were randomly matched with one of the student participants in Room Y. Your identity will be anonymous to them, and theirs to you.[30]

5. In this study you will make a single decision that may influence both your payoffs and the payoffs of the person you are matched with. Please note that there is not a correct or incorrect response, your decision is personal and yours to make.

6. The study will go as follows:

   (a) You will read the instructions which we will hand to you in a moment.

   (b) You will answer a short quiz. This quiz will be handed to you just after the instructions are finished. The purpose of the quiz is to confirm that you haver perfectly understood the instructions. Be careful, it is important that you answer the key question that you will find on the first page of the instructions *after* you have successfully completed the quiz.

   (c) Once you have completed the quiz, and after we have checked the correctness of the answers, you can answer the Key Question (the one and only real decision you will make during the course of the experiment!). Keep in mind that there is no relationship between the answers in the quiz and the answer you'll have to give the key question!

   (d) When everyone has finished we will collect your choices, leave and match your choices with the responses from Room Y, and return.

   (e) While you wait for us to calculate the amount of your winnings, we will hand out a form asking for your feedback and comments. We will also distribute a survey which is completely anonymous, but there are some personal questions so if you prefer not to respond to some of these, feel absolutely free not to.

   (f) When we are ready to proceed with payments, after collecting all of the surveys, we will call you one by one to the front of the room. You will need to bring with you the little number that we gave away at the entrance, the completed receipt form (which will be delivered towards the end of the experiment), your student ID, and all your belongings so that you can leave immediately without disturbing others.

[30]The Room X and Y designation were chosen so to make it apparent that identities would be kept anonymous. If a participant asked for more details about Room Y students, we responded to specific question, this happened twice. In reality the students in Room Y decided a few days before students in Room X decided and were paid a few days after. We did not reveal this information to keep the saliency of betrayal high.

7. We are nearly ready to have you begin reading the instructions. I would like to emphasize one more time that you read the instructions carefully. This is in your best interest because your earnings from this experiment depend largely on your our answer to the Key Question, the only decision you will make today that has monetary consequences. We remind you, please do not respond to the key question until you have completely read the instructions and responded to the brief quiz.

8. We are now ready, we will give you the instructions, and after a couple of minutes we will hand you the quizzes. Please write your code number ("little number") at the top of each sheet, so as to avoid confusion with the payments. We will now pause to answer any general questions, please raise your hand and we will come around to you individually.

9. Thank you for your attention, you may begin the instructions. If you have a question at any point, please raise your hand and we will come around to respond.

B.3 Instructions *Treatment Human (H)* (English Translation):

Welcome to the research project! Your code number is: .....
You are participating in a study in which you will earn some money. The amount you earn will depend on the outcome of a game you will play. At the end of the study, your earnings will be added to your participation fee of 5 Euros, and you will be paid in cash.

How the study is conducted. The study is conducted anonymously. Participants will be identified only by code numbers. There is no communication among them. You have been randomly paired with another participant in Room Y, call him/her "**Person Y**". Person Y will never know your identity and you will never know Person Y's identity. Your choices are identified solely by your code number and will never be disclosed to anyone. Both you and Person Y are equally informed of these instructions.

What the study is about. The study seeks to understand how people decide. You will decide between two alternatives, A and B. Alternative A gives you a certain payoff that does not depend on the choice of Y. Alternative B gives you an outcome that depends on Y's behavior. Y chooses between options **J** and **K**.

Payoff Table The payoff table reads as follows:

Result of your decision	Nature of choice	Your earnings	Earnings of Person Y
Alternative A	Certainty	10	10
Alternative B	Person Y chooses		
	Option J	15	15
	Option K	8	22

The payoff table is as follows

- If you choose A: you and Person Y will each earn 10 Euros.
- If you choose B:
  - If Person Y chooses J, you and Person Y will each earn 15 Euros.
  - If Person Y chooses K, you will earn 8 Euros and Person Y will earn 22 Euros

KEY QUESTION: For you to choose Alternative B instead of Alternative A, how large would the probability p of being paired with a Person Y who chooses option J minimally have to be? (like any probability, it must lie between 0 and 1)

YOUR ANSWER: I will choose alternative B for any p that is at least [ ] ⟵
(this means that I choose alternative A for any p that is less than this cutoff)

Note: You do not know what the actual value of p is. Your choice does not influence the value of p. It is determined by the fraction of persons Y choosing option J. With YOUR ANSWER you indicate how large the fraction of persons Y who choose J has to be before you pick B over A. This is explained in detail on the next page

Conduct of the study C.1.

1. Before knowing your choice, Person Y has to answer the following question: "Which option, J or K, do you choose in case B?" Everyone will decide in this way. After everyone has decided, we will collect the answer forms. Please fold them so that nobody can see YOUR ANSWER.
2. We will then calculate the percentage of persons Y who chose option J and inform everyone of it. This gives $p*$, the probability of being paired with a Person Y who chose option J.
3. **If $p*$ is greater than or equal to your required value of p (from YOUR ANSWER above), we will follow your instructions. Your earnings will be determined by the choice of the Person Y you are matched with.**
   (a) If your Person Y chose J, you and Person Y will earn 15 Euros each.
   (b) If your Person Y chose K, you will earn 8 Euros and Person Y will earn 22 Euros.
4. **If $p*$ is less than your required value of p (from YOUR ANSWER above), we will follow your instructions: You and your Person Y will get the outcome of the certain choice A, namely 10 Euros each.**

Completion of Study and Earnings.

- Before we conduct the study, we ask you to complete a pre-study questionnaire. We will start the study once everyone has correctly filled out this questionnaire.
- You can collect your earnings by presenting your CODE NUMBER FORM at the end of the study. Your earnings will be in an envelope marked with your code number.

B.4 Instructions *Treatment Random Device (RD)* (English Translation):

Welcome to the research project! Your code number is: .....
You are participating in a study in which you will earn some money. The amount you earn will depend on the outcome of a game you will play. At the end of the study, your earnings will be added to your participation fee of 5 Euros, and you will be paid in cash.

How the study is conducted. The study is conducted anonymously. Participants will be identified only by code numbers. There is no communication among them. You have been randomly paired with another participant in Room Y, call him/her "**Person Y**". Person Y will never know your identity and you will never know Person Y's identity. Your choices are identified solely by your code number and will never be disclosed to anyone. Both you and Person Y are equally informed of these instructions.

What the study is about. The study seeks to understand how people decide. You are confronted with two alternatives, A and B. Alternative A gives you and Person Y a payoff of 10 Euros for sure. Alternative B gives you an outcome that depends on which option (**J** or **K**) is chosen for Y. Each of the 17 cells below contains one symbol, either J or K. The symbols are visible to Person Y but not to you.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Using the online random number service www.random.org, a number between 1 and 17 will be randomly selected for Person Y. If the corresponding cell of the number selected contains J that means option J is selected for Person Y, if the cell contains K that means option K is selected for Person Y.
Payoff Table The payoff table reads as follows:

Result of your decision	Nature of choice	Your earnings	Earnings of Person Y
Alternative A	Certainty	10	10
Alternative B	Selection for Person Y		
	Option J	15	15
	Option K	8	22

The payoff table is as follows

- If you choose A: you and Person Y will each earn 10 Euros.
- If you choose B:
  - If option J is selected for Person Y, you and Person Y will each earn 15 Euros.
  - If option K is selected for Person Y, you will earn 8 Euros and Person Y will earn 22 Euros

KEY QUESTION: For you to choose Alternative B instead of Alternative A, how large would the probability p of being paired with a Person Y where option J is selected for them minimally have to be? (like any probability, it must lie between 0 and 1)

YOUR ANSWER: I will choose alternative B for any p that is at least [ ] ⟵
(this means that I choose alternative A for any p that is less than this cutoff)

Note: You do not know what the actual value of p is. Your choice does not influence the value of p. It is determined by the fraction of persons Y who have option J selected for them. With YOUR ANSWER you indicate how large the fraction of persons Y who have option J selected for them has to be before you pick B over A. This is explained in detail on the next page

Conduct of the study C.1.

1. After all the options have been selected for those in Room Y, we will first calculate the percentage of people in Room Y who have had option J selected for them and inform everyone of it. This gives $p*$, the probability of being paired with a Person Y who has had option J selected for them.
2. **If $p*$ is greater than or equal to your required value of p (from YOUR ANSWER above), we will follow your instructions. Your earnings will be determined by the option selected for the Person Y you are matched with.**
   (a) If your Person Y had option J selected for them, you and your Person Y will earn 15 Euros each.
   (b) If your Person Y had option K selected for them, you will earn 8 Euros and your Person Y will earn 22 Euros.
3. **If $p*$ is less than your required value of p (from YOUR ANSWER above), we will follow your instructions: You and your Person Y will get the outcome of the certain choice A, namely 10 Euros each.**

Completion of Study and Earnings.

- Before we conduct the study, we ask you to complete a pre-study questionnaire. We will start the study once everyone has correctly filled out this questionnaire.
- You can collect your earnings by presenting your CODE NUMBER FORM at the end of the study. Your earnings will be in an envelope marked with your code number.

B.5 Instructions *Treatment Human-Unaware (H-UA)* (English Translation):

Welcome to the research project! Your code number is: .....
You are participating in a study in which you will earn some money. The amount you earn will depend on the outcome of a game you will play. At the end of the study, your earnings will be added to your participation fee of 5 Euros, and you will be paid in cash.

How the study is conducted. The study is conducted anonymously. Participants will be identified only by code numbers. There is no communication among them. You have been randomly paired with another participant in Room Y, call him/her "**Person Y**". Person Y will never know your identity and you will never know Person Y's identity. Your choices are identified solely by your code number and will never be disclosed to anyone.

What the study is about. The study seeks to understand how people decide. You are confronted with two alternatives, A and B. Alternative A gives you a certain payoff that does not depend on the choice of Y. Alternative B gives you an outcome that depends on Y's behavior. *Persons Y are not aware they are matched with anyone or that any payoffs depend on their behavior.* Each of the 17 cells below contains one symbol, either J or K. The symbols are not visible to you or Person Y.

Without knowing the purpose, Person Y will blindly choose one of the 17 cells below. If the chosen cell contains J, that means Person Y has selected option J. If the chosen cell contains K that means Person Y has selected option K.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Using the online random number service www.random.org, a number between 1 and 17 will be randomly selected for Person Y. If the corresponding cell of the number selected contains J that means option J is selected for Person Y, if the cell contains K that means option K is selected for Person Y.
Payoff Table The payoff table reads as follows:

Result of your decision	Nature of choice	Your earnings	Earnings of Person Y
Alternative A	Certainty	10	10
Alternative B	Person Y chooses		
	Option J	15	15
	Option K	8	22

The payoff table is as follows

- If you choose A: you and Person Y will each earn 10 Euros.
- If you choose B:
  - If Person Y chooses the number that corresponds to option J, you and Person Y will each earn 15 Euros.
  - If Person Y chooses the number that corresponds to option K, you will earn 8 Euros and Person Y will earn 22 Euros

KEY QUESTION: For you to choose Alternative B instead of Alternative A, how large would the probability p of being paired with a Person Y who chooses a number that corresponds to option J minimally have to be? (like any probability, it must lie between 0 and 1)

YOUR ANSWER: I will choose alternative B for any p that is at least [ ] ⟵
(this means that I choose alternative A for any p that is less than this cutoff)

Note: You do not know what the actual value of p is. Your choice does not influence the value of p. It is determined by the fraction of persons Y who chose a number corresponding to option J. With YOUR ANSWER you indicate how large the fraction of persons Y who chose a number corresponding to option J has to be before you pick B over A. This is explained in detail on the next page

Conduct of the study C.1.

1. After all everyone has made their decision, we will first calculate the percentage of people in Room Y who have chosen a number corresponding to option J and inform everyone in Room X of it. This gives $p*$, the probability of being paired with a Person Y who has chosen a number corresponding to option J.

2. **If $p*$ is greater than or equal to your required value of p (from YOUR ANSWER above), we will follow your instructions. Your earnings will be determined by the the choice of the Person Y you are matched with.**

   (a) If your Person Y has chosen a number corresponding to option J, you and your Person Y will earn 15 Euros each.

   (b) If your Person Y has chosen a number corresponding to option K, you will earn 8 Euros and your Person Y will earn 22 Euros.

3. **If $p*$ is less than your required value of p (from YOUR ANSWER above), we will follow your instructions: You and your Person Y will get the outcome of the certain choice A, namely 10 Euros each.**

Completion of Study and Earnings.

- Before we conduct the study, we ask you to complete a pre-study questionnaire. We will start the study once everyone has correctly filled out this questionnaire.
- You can collect your earnings by presenting your CODE NUMBER FORM at the end of the study. Your earnings will be in an envelope marked with your code number.

B.6 Instructions *Treatment Human-Unforeseeable (H-UF)* (English Translation):

Welcome to the research project! Your code number is: .....
You are participating in a study in which you will earn some money. The amount you earn will depend on the outcome of a game you will play. At the end of the study, your earnings will be added to your participation fee of 5 Euros, and you will be paid in cash.

How the study is conducted. The study is conducted anonymously. Participants will be identified only by code numbers. There is no communication among them. You have been randomly paired with another participant in Room Y, call him/her "**Person Y**". Person Y will never know your identity and you will never know Person Y's identity. Your choices are identified solely by your code number and will never be disclosed to anyone. Both you and Person Y are equally informed of these instructions.

What the study is about. The study seeks to understand how people decide. You are confronted with two alternatives, A and B. Alternative A gives you a certain payoff that does not depend on the choice of Y. Alternative B gives you an outcome that depends on Y's behavior. Each of the 17 cells below contains one symbol, either **J** or **K**. The symbols are not visible to you or Person Y.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Person Y will blindly choose one of the 17 cells. If the chosen cell contains J, that means Person Y has selected option J. If the chosen cell contains K that means Person Y has selected option K. This means that regardless of which option Person Y prefers, the option will be selected only if Person Y's number choice yields that option.
Payoff Table The payoff table reads as follows:

Result of your decision	Nature of choice	Your earnings	Earnings of Person Y
Alternative A	Certainty	10	10
Alternative B	Person Y chooses		
	Option J	15	15
	Option K	8	22

The payoff table is as follows

- If you choose A: you and Person Y will each earn 10 Euros.
- If you choose B:
  - If Person Y chooses the number that corresponds to option J, you and Person Y will each earn 15 Euros.
  - If Person Y chooses the number that corresponds to option K, you will earn 8 Euros and Person Y will earn 22 Euros

KEY QUESTION: For you to choose Alternative B instead of Alternative A, how large would the probability p of being paired with a Person Y who chooses a number that corresponds to option J minimally have to be? (like any probability, it must lie between 0 and 1)

YOUR ANSWER: I will choose alternative B for any p that is at least [ ] ⟵
(this means that I choose alternative A for any p that is less than this cutoff)

Note: You do not know what the actual value of p is. Your choice does not influence the value of p. It is determined by the fraction of persons Y who chose a number corresponding to option J. With YOUR ANSWER you indicate how large the fraction of persons Y who chose a number corresponding to option J has to be before you pick B over A. This is explained in detail on the next page

Conduct of the study C.1.

1. After all everyone has made their decision, we will first calculate the percentage of people in Room Y who have chosen a number corresponding to option J and inform everyone in Room X of it. This gives $p*$, the probability of being paired with a Person Y who has chosen a number corresponding to option J.
2. **If $p*$ is greater than or equal to your required value of p (from YOUR ANSWER above), we will follow your instructions. Your earnings will be determined by the the choice of the Person Y you are matched with.**
   (a) If your Person Y has chosen a number corresponding to option J, you and your Person Y will earn 15 Euros each.
   (b) If your Person Y has chosen a number corresponding to option K, you will earn 8 Euros and your Person Y will earn 22 Euros.
3. **If $p*$ is less than your required value of p (from YOUR ANSWER above), we will follow your instructions: You and your Person Y will get the outcome of the certain choice A, namely 10 Euros each.**

Completion of Study and Earnings.

- Before we conduct the study, we ask you to complete a pre-study questionnaire. We will start the study once everyone has correctly filled out this questionnaire.
- You can collect your earnings by presenting your CODE NUMBER FORM at the end of the study. Your earnings will be in an envelope marked with your code number.