Social Preference Reversals

Testing a common account of choice under uncertainty, intertemporal choice and interpersonal choice

Geoffrey Castillo[*], Nick Netzer[**], Daniel Seidmann[*], and Chris Starmer[*]

[*]University of Nottingham
[**]University of Zurich

14th February 2014

Abstract

Previous research suggest that choice under uncertainty, intertemporal choice and interpersonal choice might be linked, yet the risk, time and social dimensions are treated separately in traditional models. We propose a common account of those choices. We treat probabilities, delays and social distances as a common psychological distance. Self-control is a potential candidate by which this distance is mediated in choice. If this interpretation is correct, classical anomalies observed in the context of choice under uncertainty and intertemporal choice should also be observed in the context of interpersonal choice. We focus on the standard preference reversal and hunt its social analogue – hence the social preference reversal.

Contents

1 Introduction

2 Motivation

2.1 Similarities between choice under uncertainty and intertemporal choice

The traditional models of choice under uncertainty and intertemporal choice do not allow for any interaction between the risk and time dimensions. The expected utility model is linear in probabilities and captures attitude to risk with the utility function, and the discounted utility model captures attitude to time with the discount function and the pure rate of time preference. When those two are combined into the discounted expected utility model, risk and time are strictly segregated and there is no possibility for uncertainty to affect the discount rate or for delay to affect the attitude to risk. Even recent models such as prospect theory or hyperbolic discounting, more behaviourally founded, reproduce this clean distinction.

There are, however, studies showing that choice under uncertainty and intertemporal choice might be linked. Rachlin et al. (1986) were the first to explicitly mention and investigate this possibility. They draw a parallel between experiments on probability weighting in humans (Kahneman and Tversky, 1979) and experiments on delay discounting in animals, e.g. pigeons (Rachlin and Green, 1972), and they show that there is a one-to-one mapping between the two sets of experimental findings. They hypothesise that a risk preference is essentially a time preference with the following explanation: the less probable is an event, the more one has to wait to see its realisation. They test and verify this idea in an experiment in which subjects become more risk averse as the delay to see the gamble realised increase. Rachlin pursued in this direction with Raineri and Cross (1991) and with Siegel (1994) and added that probability weighting functions and delay discount functions are both hyperbolic and very similar for a given subject.

This interpretation, that delay discounting is the fundamental process behind uncertainty, is promoted mostly by Rachlin and colleagues. The reverse causality, that a delay is inherently an additional risk and risk preference is the fundamental process, is supported by the remaining of the experiments in this line of research (Benzion et al., 1989; Keren and Roelofsma, 1995; Weber and Chapman, 2005; Baucells and Heukamp, 2010). For example, Keren and Roelofsma (1995) and Weber and Chapman (2005) show that adding uncertainty counters the immediacy effect, and that an additional delay produces effects similar to the ones obtained when uncertainty is increased.

This has soon triggered theoretical investigation. Prelec and Loewenstein (1991) started by showing that the expected utility model and the discounted utility model can be derived from almost the same set of axioms. There is indeed nothing empirical in

this proof, but to the extent that both models try to capture something that people do when dealing with probabilities or delays, this indicates that there could be common processes at play. Quiggin and Horowitz (1995) use the same type of argument but using rank-dependent expected utility. Halevy (2008) –in a paper with a very well-chosen title: 'Strotz meets Allais'– went one step further by building a model where the delay discount function is essentially a probability discount function. This is done following the explanation given above: delay is interpreted as an additional uncertainty. Interestingly, he shows that in this framework, using a non-linear probability weighting function à la Kahneman and Tversky (1979), the delay discount function generates reversals, i.e. it is hyperbolic. Epper et al. (2011), albeit having a different framework, presents almost the same result: uncertainty generates an hyperbolic discount function. This research agenda culminated with Baucells and Heukamp (2012) who derived a theory of choice under risk and intertemporal choice from the set of axiom used by Fishburn and Rubinstein (1982), a theory in which the decision-maker makes trade-off between risk and time dimensions.

Even if the dispute between the proponents of the two types of causality cannot be settled, there is evidence that choice under risk and intertemporal choice are in some ways intertwined. If it was needed, further confirmation is added by Todorov et al. (2007) who apply construal-level theory from social psychology to show that probability and delay are similar psychological distances.

Thus, risk and time seem to be linked. Interestingly enough, similar studies and similar conclusions have been reached for intertemporal choice and interpersonal choice.

2.2 Similarities between intertemporal choice and interpersonal choice

Albert (1977) was the first to draw a parallel between temporal behaviour and social behaviour. Strikingly, his way of proceeding is not very different from Prelec and Loewenstein (1991). He starts from Festinger's (1954) social comparison theory and he shows that it can be translated into a theory of intertemporal choice with a very small set of rewriting rules. Those rewriting rules state how a proposition or a word in one theory should be translated into the other theory. Again, this is a purely theoretical exercise but it strongly suggests that time and social domains might have some links.

This idea convinced Rachlin and Jones (2008) to investigate the relations between delay discounting and social discounting. Social discounting (Jones and Rachlin, 2006) refers to the decrease in the amount one is willing to give to someone as this someone is more socially distant from her. Their way of eliciting it was to ask subjects to imagine they could classify their acquaintance in a list from 1 to 100, where 1 is the closest and 100 is the more distant. Then, they asked subjects to choose between: A) $\$x$ for you alone; B) \$75 for you and \$75 for the n^{th} person on the list. By varying x and n and

computing switching points, they uncovered what they called a social discount function with a hyperbolic property. Not only both delay discount functions and social discount functions have the same hyperbolic form, Rachlin and Jones also observed that anomalies in the time domain, for example the amount effect which makes the discount function steeper as amounts increase, are also present in the social domain with the social discount function. They were even able to predict subjects' behaviour in social contexts solely with the estimation of their delay discount function.

Pronin et al. (2008) propose that this happens because people treat their future selves as different people. In other words, we experience social distance and temporal distance in a similar way. They test this idea by showing that people make similar choices for their future selves and for other people, and that those choices are very different from the choices made for the present self. Their explanation of this phenomenon builds on Pronin and Ross (2006): we pay less attention to the subjective experience of an event when dealing with future selves or other people; conversely, when dealing with decisions affecting the present self, the emphasis is put on subjective experiences.

Thus, after having presented parallels between risk and time dimensions, we see that there are also parallels between time and social dimensions. There is then a natural question to ask: is there some kind of transitivity at play?

2.3 Self-control: a common account for the choices across the three domains?

If choice under uncertainty and intertemporal choice are linked, and if intertemporal choice and interpersonal choice are also linked, then could the three contexts be somehow related?

This has been in the air for quite some time and can be read between the lines of the research presented above. For example, Baucells and Heukamp (2012) use a 'psychological distance function' which can take as inputs 'risk distances' or 'time distances'. The psychological distance is also at the heart of constual-level theory (Trope and Liberman, 2010). This theory speculates that the more psychologically distant is an object, the more abstract it is mentally represented. The distance is taken between the present self and the object, and the object can be almost anything: something in the future or in the past, a gamble or even another person. Thus, the psychological distance in construal-level theory covers the three contexts: risk, time and social.

Even without this evidence, choice under uncertainty, intertemporal choice and interpersonal choice look strikingly similar: in the three cases, outcomes have a concrete dimension, e.g. money; the second dimension, be it a delay, a probability or the closeness of a relationship, can be interpreted as a general 'distance'.

Let me be more precise. The decision-maker deals with bi-dimensional objects of the

Table 1: Domains.

	Risk	Time	Social
Distance d	Probability p	Delay t	Closeness s
Object of choice (x, d)	Gambles (x, p)	Dated outcome (x, t)	Allocations (x, s)

form (x, d), where x is for simplicity an amount of money and d a distance. In the case where d is a probability, the objects are (x, p), that is, gambles. In the case where d is a delay, the decision-maker faces dated outcomes (x, t). Finally, when the distance is a social distance, e.g. the closeness of a relationship, we have allocations (x, s). These three domains of decision-making are summarised in Table 1. Delay and social distance are already standard distances. Probabilities are a bit different because they only cover the interval $[0, 1]$. To remedy to this problem, we could use the transformation $d = -\ln p$.

The natural way to think about the mean by which people deal with temporal distances, i.e. delays, is to think about self-control. People choose smaller-sooner rewards over larger-later rewards because of impatience and lack of self-control. Those with more willpower are more likely to resist the temptation of the soonest reward. This oversensitivity to immediacy is also observed in the case of probabilities, where people tend to choose near-certain gambles over uncertain ones. Again, it is as if people could not resist the temptation of high probabilities. Generally, in both cases, people cannot resist the attraction of small distances.

We propose to extent this reasoning to allocations. The direct consequence of this is that *altruism and self-control might be linked*. This the conclusion also reached by Kanfer (1979) in his attempt to find a mechanism able to promote altruism. He sees altruism as the result of two forces: external control, e.g. religion or society; and internal control, e.g. one's own morality or willpower. He considers that high levels of cooperation among human groups cannot be sustained through traditional forms of external control any more, and that new forms of external control are hard to find. His answer to this conundrum is the promotion of internal control. The reason is that altruistic acts involve a conflict that is also found in situations where an individual has to exert self-control. As he puts it, 'it is proposed that the initial task for training altruistic behavior consists of helping a person to alter the probability of the highly preferred and generally more immediately rewarded response (one that is egocentrically oriented) in favor of a response that has an initially lower probability because it is less immediately rewarding.' This is similar to the situations described earlier where people prefer sooner-lower rewards to later-larger rewards.

Several studies support this view. For example, Curry et al. (2008) find a negative correlation between discount rates and contributions in a public good game. Like Kanfer (1979), their interpretation is that for altruism to emerge, you need to be patient and to

exert self-control.

Martinsson et al. (2010b) investigate whether acting pro-socially or selfishly is linked to self-control. In their understanding, basic urges for personal gain and greed tempt the decision-maker and enter in conflict with the willingness to cooperate and to act pro-socially. They use a framework developed by Myrseth and Fishbach (2009) which sees self-control as a 2-stage process: first, the self-control conflict must be identified, and second, self-control must be effectively applied. The basic idea of their experiment is to manipulate the degree of identification of self-control and to see its impact on altruism in a linear public good game and a dictator game. For the manipulation, they follow Myrseth and Fishbach (2009) and they show a calendar to the subjects before they play the game. In the isolated frame treatment, the calendar has a grid separating the days of the month. In the interrelated frame treatment, there is also a calendar view of the current month but no grid. As a control group, they have a standard frame treatment without calendar. Their idea is that in the isolated frame treatment, seeing the grid makes the subject less likely to identify self-control conflicts because it nudges them to bracket narrowly. In contrast, in the interrelated frame treatment, seeing a calendar without grid makes them more likely to frame the problem broadly and to identify self-control problems. However, it should be noted that the outcomes in both games are not dated. Hence, the calendar is only supposed to nudge the subjects to treat the problem in a particular way without having any actual purpose for the experiment. Finally, they measure the ability to exert self-control using the Rosenbaum scale, which measures the capacity to exert self-control in the face of temptation. Their results show that in both game there is a correlation between pro-social behaviour and self-control ability in the interrelated frame treatment, but this correlation disappears in the isolated frame treatment. In other words: even if you can exert self-control, you need to identify the fact that you face a self-control conflict in order for self-control to have some bite. Martinsson et al. (2010a) use the same principle to explain imperfect conditional cooperation. Conditional cooperation refers to the fact that, in a public good game, you contribute as long as the others contribute, which you observe using the strategy method. Their idea is that a lack of willpower may impede perfect conditional cooperation, which explains why we generally observe imperfect conditional cooperation. Kocher et al. (2012) is basically again the same story but with a formal model that incorporates other-regarding preferences and a cost of self-control.

Achtziger et al. (2011) add ego depletion to the picture. The ego depletion hypothesis (Baumeister et al., 1998) states that volition is a limited resource and that making active choice depletes this resource. The more the resource is depleted, the harder it is for the decision-maker to make active choices and to counter automatic and implicit processes. In the case of interpersonal choice, the authors assume that acting pro-socially is an explicit and rational process that requires volition to be implemented. On the contrary, acting selfishly is an implicit process and an urge. Putting everything together, they hypothesise

Table 2: Phenomena.

Risk	Time	Social
Common difference effect:		
Common ratio	Common delay	Common social distance
Choice/valuation preference reversal:		
Choice/valuation of gambles	Choice/valuation of delayed reward	Choice/valuation of allocations

that the more depleted is the decision-maker, the less he will act pro-socially because he will rely on the automatic choice and his temptation to be selfish. Indeed, in a standard ultimatum game, they observe that depleted proposers make lower offers, because they are less able to control their impulse of keeping the whole pie for them. Similarly, they observe that depleted responders reject more often unfair offers, because it requires self-control to counter the urge of punishing the proposer.

To summarise, we think that the same process might govern choice under uncertainty, intertemporal choice and interpersonal choice. In the three cases, the decision-maker has to deal with objects of the form (x, d), d being a psychological distance. Self-control seems to be a good candidate for such a process. However, this idea would be entirely void if we would not propose a possible way of testing it.

2.4 The social preference reversal

If the same process is behind the choices made in the three domains, then we should observe similar patterns of choice and similar anomalies.

I concentrate on two of the most-observed phenomena in experiments: *common difference effects* and *standard preference reversals*. *Common difference effects* (first row of Table 2) happen when $(x, d) \succ (x', d')$ but $(x', d' + \delta) \succ (x, d + \delta)$. One readily recognises the common delay effect in intertemporal choice. The common ratio effect is not that different and the similarity between the two effects as already been suggested in previous research (for example in Prelec and Loewenstein, 1991). The common ratio effect happens when $(x, p) \succ (x', p')$ but $(x', p' \times \pi) \succ (x, p \times \pi)$ and the similarity becomes clear when one applies the transformation $d = -\ln p$. The literature on common difference effects in the three domains is reviewed in the first row of Table 3. The majority of the studies I presented in Sections 2.1 and 2.2 use this type of problems. For example, Weber and Chapman (2005) shows that common ratio effects can be generated using delays and that common delay effects can be generated using probability ratios. Jones and Rachlin (2006) use common difference problems in the social domains, and find a 'common social distance' effect. Hence, parallels between the three types of common difference effects

Table 3: Literature.

Risk	Time	Social
Common difference effect:		
Kahneman and Tversky (1979), Starmer and Sugden (1989), Cubitt et al. (1998)	Thaler (1981), Benzion et al. (1989), Green et al. (1994), Kirby and Herrnstein (1995), Keren and Roelofsma (1995), Kirby and Maraković (1995)	Jones and Rachlin (2006),Rachlin and Jones (2008), Jones and Rachlin (2009)
Chapman and Weber (2006) Baucells and Heukamp (2010)		
Choice/valuation preference reversal:		
Lindman (1971), Lichtenstein and Slovic (1971), Lichtenstein and Slovic (1973), Grether and Plott (1979), Cubitt et al. (2004)		?
Tversky et al. (1990)		

have already been stressed.

Standard preference reversals, or *choice/valuation preference reversals* in order to avoid confusion[1] (second row of Table 2), have mostly been studied in the context of choice under uncertainty. It occurs when people deal with \$-bets and P-bets, \$-bets being rich in the money dimension and P-bets in the probability dimension. With notations $(x^{\$}, p^{\$})$ and (x^P, p^P), we have $x^{\$} > x^P$ and $p^P > p^{\$}$. It is observed that in choice tasks subjects reveal $(x^P, p^P) \succ (x^{\$}, p^{\$})$ but in valuation tasks they reveal $V\Big((x^{\$}, p^{\$})\Big) > V\Big((x^P, p^P)\Big)$, making it a preference reversal. The second row of Table 3 lists the main papers on this subject. Interestingly, this shows that only Tversky et al. (1990) explicitly study the choice/valuation preference reversal in the context of intertemporal choice. In that case, subjects choose sooner-lower options but value more the later-larger options.

In Section 2.3, we claimed that choices in the three domains might be combined into an all-encompassing model in which the distance to the consequence matters. This means that we could redefine the choice/valuation preference reversals in those terms to accommodate the three instances of the choice/valuation preference reversal presented in the second row of Table 2. Hence, such a reversal is likely to occur when people deal with \$-options, with a high but distant reward, and C-options – close options – with a low but close reward. With notations $(x^{\$}, d^{\$})$ and (x^C, d^C), we have $x^{\$} > x^C$ and $d^{\$} > d^C$, making it a trade-off. In choice tasks, we have $(x^C, d^C) \succ (x^{\$}, d^{\$})$, but in valuation tasks $V\Big((x^{\$}, d^{\$})\Big) > V\Big((x^C, d^C)\Big)$. The reversals mentioned in the previous paragraph can be easily recovered if one replaces d by probabilities p or delays t.

If the interpretation of probability, delay and social distance as a unique psychological distance is correct, then it means that we *should* observe a choice/valuation preference reversal in the social domain, as it is observed in the risk and the time domain. Hence, our way of testing the relevance of the psychological distance interpretation is to hunt the *social preference reversal*.

To conclude, we propose to treat risk, time and social distance as a unique psychological distance. This is supported by studies investigating common difference effects across the three domains, and by the observation that choice under uncertainty, intertemporal choice and interpersonal choice look structurally similar. In order to test this idea, we extent the choice/valuation preference reversal to the social domain. The next section develops this idea, introduces the research question and presents an experimental design aimed at hunting the social preference reversal.

[1]The terms 'preference reversal' are sometimes used for common difference effects.

3 Experimental design: theory...

3.1 Research question

Through our experiment, we aim to answer two questions:

1. Is there a choice/valuation preference reversal in social domains, that is, a social preference reversal?

2. Are reversals across the three domains related?

For the time being and for the pilot experiment, we focus on question 1. The reason is simple: if there is no such thing as a social preference reversal, question 2 makes no sense. Hence, in a first time, we hunt the preference reversal in the context of interpersonal choice. In a second time, if the answer to question 1 is positive, then we will focus on investigating preference reversals in the three domains and look at a particular subject's behaviour across risk, time and social contexts.

I will now present what an experiment aimed at answering the research question 1 might in principle look like. Essentially, we translate standard preference reversal experiments to the social domain, the objects of choice being now allocations and not gambles or dated rewards. The translation is straightforward except for two tedious elements: valuation tasks and social distances.

3.2 Setup

We previously defined \$-options and C-options which can fit any of the three domains. We now focus on the social domain.

We denote by (x, s) allocations providing an amount of money x to someone with a social distance s between himself or herself and the decision-maker. We will use a slight abuse of language and use s both for the social distance itself and for the person situation at a social distance s. For example, the statement 'allocating x to s' is a short form of 'allocating x to someone socially distant of s from the decision-maker.' We are specifically interested in \$-allocations $(x^{\$}, s^{\$})$ and C-allocations (x^{C}, s^{C}). We impose the conditions $x^{\$} > x^{C}$ and $s^{C} > s^{\$}$: \$-allocations provide a large amount of money to someone socially distant to the decision-maker and C-allocations provide a small amount of money to someone socially close.

I will deliberately remain vague on social distances and what s really is. The reason is that it will mainly be dictated by the way we choose to implement the experimental design. For example, s can be the degree of friendship if we use subjects that know each other in real life, or the indication of belonging to the same group as the decision-maker. Thus, being precise on social distance is relegated to the next section which focuses on

implementation. For the time being, please bear with me and adopt a loose, unspecified definition of social distance – for example: how close you are to someone.

3.3 Tasks

As in standard preference reversal experiments, subjects undertake two tasks: choice tasks and valuation tasks.

Choice tasks simply ask the subjects to choose between a \$-allocation and a C-allocation. The questions are of the form: 'do you prefer to give $x^{\$}$ to $s^{\$}$ or x^{C} to s^{C}?' I denote by '$\succ$' the preference revealed in the choice task. Incentives are straightforward: the allocation chosen by the subject is realised. In the case of several $x - s$ combinations and several choices made by a subject, one choice is selected at random and played for real.

Valuation tasks are in principle also straightforward. A valuation of an allocation (x, s) is an amount m such that $(x, s) \sim (m, s')$. m being the valuation of (x, s), we can write $m = V\Big((x, s)\Big)$. The real question is: what should be s'? In the context of choice under uncertainty, the valuation is a certainty equivalent: $(x, p) \sim (m, 1)$. In the context of intertemporal choice, the valuation is an immediacy equivalent: $(x, t) \sim (m, 0)$. The logical translation for interpersonal choice is a self equivalent: $(x, s) \sim (m, 0)$; or in words: m such that there is indifference between x given to s and m given to the decision-maker. Questions would be of the form: 'state m that makes you indifferent between giving x to s and m to yourself.' However, this is not exactly the social equivalent of risk and time valuations. Valuations in the risk and time domains involve only the self: you are given, say, a gamble, and you are asked to state a minimum selling price. The valuation task for the social domain just presented does not involve only the self because it requires you to value an amount of money given to someone else. Hence, this can potentially create confounds. To address this problem, other possibilities exist. In principle, s' can be anything as soon as it stays the same across all choices made by a given subject in order to allow for comparison of the different valuations. Hence, s' can be a reference person with a social distance $s' \neq s$. s' can also be anything easily identifiable for a subject, such as a charity. Incentives in the valuation tasks could mirror the ones typically used the domain of risk with a mechanism à la Becker, DeGroot, and Marschak (1964).

3.4 Hypotheses

Now that we have set up notations and tasks to be used, we can state the following hypotheses:

H_0 There is no social preference reversal. For a given subject, one of the two following patterns of behaviour is observed:

- $(x^\$, s^\$) \succ (x^C, s^C)$ and $V\Big((x^\$, s^\$)\Big) > V\Big((x^C, s^C)\Big)$
- $(x^C, s^C) \succ (x^\$, s^\$)$ and $V\Big((x^C, s^C)\Big) > V\Big((x^\$, s^\$)\Big)$

H_a There is a social preference reversal. There are two sub-cases:

H'_a *Strong social preference reversal*: for a given subject, we observe $(x^C, s^C) \succ (x^\$, s^\$)$ but $V\Big((x^\$, s^\$)\Big) > V\Big((x^C, s^C)\Big)$.

H''_a *Weak social preference reversal*: the proportion of

$$\Big\{(x^C, s^C) \succ (x^\$, s^\$) \text{ but } V\Big((x^\$, s^\$)\Big) > V\Big((x^C, s^C)\Big)\Big\}$$

is greater than

$$\Big\{(x^\$, s^\$) \succ (x^C, s^C) \text{ but } V\Big((x^C, s^C)\Big) > V\Big((x^\$, s^\$)\Big)\Big\}$$

If the experimental data rejects H_0 and does not reject H_a, this would be evidence that the risk, the time and the social domains might be deeply linked. However, for the time being, the experiment would not be able to discriminate between different possible sources of this linkage. We proposed earlier an explanation based on self-control, but an explanation using the scale compatibility hypothesis (Slovic et al., 1990) could also well explain the acceptance of H_a. An experiment precisely designed to distinguish between the self-control hypothesis and, for example, the scale compatibility hypothesis, is an interesting route to follow in the future.

The rejection of H_a would also be an interesting outcome. It would mean that the three domains are not as intertwined as we initially thought. It would call into question a part of the literature reviewed earlier which advocates for a common account for the choices made in the three situations. It would also allow to see more clearly what is similar and what is fundamentally different. A possibility is that risk and time and perceived similarly, but that social choice is governed by separate cognitive processes.

4 . . . and practice: Experimental implementation

We now implement the experimental design presented above. Its core elements are: \$-allocations, C-allocations, choice tasks and valuation tasks. Operationalising those elements boils down to specify what is precisely 'social distance'. Different definitions of s will trigger different recruitment processes and different procedures.

4.1 What is a social distance?

4.1.1 Definition

Loosely speaking, social distance refers to the degree of closeness between two individuals. This definition quickly reaches its limits because what is 'closeness' then? One has to dig deeper in social distance.

Social distance can be thought as having four dimensions: affective, normative, interactive and habitual (Karakayali, 2009). The affective dimension is probably the one that has been recognised the first, starting with Bogardus (1925) and its social distance scale. In Bogardus' acceptance, we are close to the people we *feel* close to. Putting it differently, we are socially close to the people we have empathy for. The normative dimension is linked to social classes: a norm defines who can de declared as 'us' and who are 'them'. In this case, social distance overlaps with group membership. The interactive dimension is the most objective one as it refers to the frequency and the length of interactions. This is the definition typically found in the network literature. Finally, the habitual dimension refers to cultural similarities. The more two group have common cultural references and the more they can be said to be socially close on the habitual dimension. In Bourdieu's terminology, it can be said that they share cultural (and possibly symbolic) capital. Indeed, the four dimensions can be more or less active at a given time and interact with each other.

These four dimensions only partially overlap. For example, you see some colleagues every day at work but it cannot be said that you consider yourself socially close to them. Conversely, you may not see your best friend very often, but still you consider him to be socially close. This shows that there can be a struggle between the affective and the interactive dimension. There is a similar problem between the affective and the habitual dimension: it is not because you share a lot of cultural elements with someone that you will feel closer to him or her. It is maybe more likely, but not sure.

Depending on the definition of social distance we adopt, the experimental procedure has to be changed to reflect the dimension favoured. An experiment based on social distance as affective closeness would be different from an experiment on social distance as interactions. This is illustrated by the way social distance has been used in the experimental literature.

4.1.2 Social distance in the laboratory

'Social distance' has had different meanings depending on the context. Hence, as I discuss the papers, I will try to make clear what was precisely meant by 'social distance'. This review will give some insights on how to operationalise this notion in our own social preference reversal experiment.

Before proceeding, a word of caution: I am only interested in the way these studies

define and implement social distance and not so much on their findings, for example the influence of social distance on cooperation. That is because they have used well-known games (ultimatum games, dictator games, trust games, public good games, prisoners' dilemmas) presenting self vs. others allocations while we are interested in others vs. others allocations for the social preference reversal. In other words, research reviewed below mostly investigates conflicts between selfishness and altruism, which is not part of our research question. For example, in our design, there is no pie to be shared between the decision-maker and a recipient and there is no real dictator.

Keeping this disclaimer in mind, it turns out that social distance has been used in four different ways.

Anonymity or identification? Hoffman et al. (1994, 1996) were probably the first to introduce social distances in an experiment and to investigate its impact on altruism in dictator games. In their words, '[s]ocial distance can be defined as the degree of reciprocity that subjects believe exist within a social interaction.' More or less reciprocity means larger or lower social distances. Hence, in order to vary social distance, the authors compare single-blind and double-blind procedures. In the former, there is anonymity between the subjects but not between the subjects and the experimenters; in other words, the experimenter learns subjects' actions and can map an action to a particular subject. In the latter, there is also anonymity between the subjects and the experimenter. They essentially show that the amount transferred by the dictator decreases as isolation increases. Hence, the higher the social distance, the lower the altruism.

This has spanned numerous studies investigating the impact of social distance as anonymity on other-regarding preferences (Bohnet and Frey, 1999a,b; Bolton et al., 1998; Charness and Gneezy, 2008; Dufwenberg and Muren, 2006; Frohlich et al., 2001, 2004; Johannesson and Persson, 2000). The key message of this line of research is that social distance is more than simply anonymity and reciprocity. Bohnet and Frey (1999a), in a direct answer to Hoffman et al. (1996), showed that 'social distance influences other-regardedness independent of any norm or social exchange.' They still use the double-blind procedure but they vary whether or not the dictator can identify his recipient. Here, more identification implies lower social distances. They find that the amount given to recipient increases as identification increases. Hence, the source of the increase in contribution when the social distance decreases does not seem to come from reciprocity motivations alone but also from identification. This is also the conclusion reached by Charness and Gneezy (2008): knowing the name of the recipient increases the amount of money sent, even when dictators and recipients are drawn from different universities.

Dufwenberg and Muren (2006) found a similar result. They use a standard dictator game with two payment conditions: private payment and payment on stage. The former is the standard procedure; in the latter, the dictator receives his payment in front of

an audience in an auditorium. If anonymity is a synonym of social distance and if it is the main mechanism behind positive donations in the dictator game, we should observe more giving in the payment on stage condition than in the private payment condition. Dufwenberg and Muren obtain precisely the opposite result: dictators give less in the on-stage treatment than in the private treatment. This makes them conclude that 'it is problematic to organize experimental data in terms of social distance if this notion is taken to vary one-to-one with anonymity. As anonymity changes other things may change alongside so that confounding factors may inadvertently be introduced.'

The debates does not seem to be settled, but at this stage what we can say is that anonymity is an obvious candidate for social distance, but it should not be regarded as the only possibility. The possibility to identify the recipient is also a strong one.

Characteristics in common. Social distance can also be measured in terms of the number of characteristics in common between two persons. This is maybe the most obvious candidate when one does not want to use anonymity.

Meer and Rigbi (2011) use an online microfinance platform to investigate the determinants of altruism. Messages on this platform are sent in different languages. They hypothesise that the language of the message is a strong determinant of social distance. The social distance is small when the sender write in the same language as the receiver, and it is large when the message is written in a different language. This is a binary social distance: $s = 0$ when both parties have a common language; $s = 1$ when they have not. Etang et al. (2011) use the same binary social distance. In a field study in Cameroon, they consider there is a large social distance between two persons not living in the same village, and a small social distance between two persons living in the same village. Again, $s = 0$ when in the same village and $s = 1$ when not living in the same village.

The characteristics in common is one of the dimensions of social distance investigated by Charness et al. (2003). They compare a laboratory experiment and an internet experiment. They assume that social distances are lower in the laboratory for three reasons. First, subjects can see each other – this is identification à la Bohnet and Frey (1999a). Second, they can be friends or acquaintances – something I will come back to later. Third, subjects being students, they share common traits such as school, age group, nationality...

Glaeser et al. (2000) also find that social distance can be expressed in terms of characteristics in common, but this is a conclusion of their study and not an hypothesis. They find that national and racial differences are the strongest predictors of trust. One can say that $s = 0$ when race and nationality are different, $s = 1$ when they are the same, and $s = 1/2$ when only one is in common.

Incidentaly, the construal-level theory I already mentioned earlier has also interpreted social distance as 'interpersonal similarity' in Liviatan et al. (2008).

Social identity. The social identity literature posits that we derive part of our sense of identity of a group. Research in social psychology starting from Tajfel et al. (1971) has shown that groups mean something to people and their group identity affects their behaviour.

Two methods have been used to create group identities: priming of natural social identities and induced group identities. The priming of natural social identities is not far from the idea of social distance as natural characteristics. Having some characteristics in common with someone is already being in the same 'natural group'. Priming means that the experimenter makes this more salient. For example, one may create a questionnaire focusing on the subject's nationality before presenting him allocations to people with different nationalities. This way, the experimenter primes a stimulus that will be showed again at a later stage.

The second method is to induce group identity. More particularly, the main method used to test the social identity theory has been the minimal group paradigm (Tajfel and Turner, 1979). The idea is to assign subjects to different groups on the basis of irrelevant elements or meaningless tasks. In this case, there is a small social distance towards ingroup members and a large social distance toward outgroup members. The classical example in this literature is to ask subjects to choose between a painting by Klee and a painting by Kandinsky. It is then observed that such a trivial element is sufficient to create an ingroup bias, i.e. favouritism toward ingroup members. Tajfel and Turner (1979) identify four conditions for a group to be a minimal group: subjects have to be assigned randomly to different groups on the basis of some trivial tasks; there is no social interaction between groups; the group membership is anonymous; and finally, there is no link between self-interest and ingroup favouritism. This fourth condition is often not verified in economic experiments because games of the like of dictator games or trust games have indeed an element of self-interest. However, this condition is satisfied in our design because, as already stressed earlier, we are interested in other vs. other allocations and not in self vs. other allocation, and thus self-interest is not part of our design – at least in the choice tasks.

Even without satisfying this fourth condition, the minimal group paradigm has been implemented several times in economics. Chen and Li (2009) create group memberships using the Klee vs. Kandinsky question and they show that subjects are more altruistic towards ingroup members in a variety of games. Eckel and Grossman (2005) look at the impact of manufactured team identity on cooperation. There are three levels of team identity: no team identity, where subjects are randomly assigned to groups with no team identification; weak group identity, where there is team identification; and strong group identity, where there are also team goals and an ingroup/outgroup conflict. They show that team membership affects cooperation only when the salience of the team identity is increased with preproduction exercises taking place within the group. This is also the

conclusion reached by Charness et al. (2007). They find that group membership in itself is not sufficient to change behaviour, but as soon as the salience of the group is increased then the ingroup bias is recovered.

It is not clear why minimal groups are sometimes not enough in economic experiments to change behaviour. This might be linked to the fourth condition for a group to be minimal typically not being satisfied in economics as soon as a game is used. This loophole does not apply to our setup, hence minimal groups might be able to create social distances.

Networks. The fourth way to think about social distance is to think about networks. This is best illustrated by an example. Goeree et al. (2010) elicited the network of friends of students in grade 5 through 12 in an all-girl school in California. To do so, the students where asked to name up to five friends and to specify the time they spent which each of them. They were also asked to state how much time they spent which other friends. Then, a subsample of all of the students who participated in the network elicitation questionnaire participated in dictator games where the social distance between the dictator and the recipient was varied. More specifically, the authors used the following social distances: $s = 1$ for immediate friends, $s = 2$ for friends of friends, and $s = 3$ and higher for more distant acquaintances. They observed that the amount a dictator is willing to give to a recipient has an inverse relationship with the social distance. More specifically, it follows a $1/s$ law. This is very interesting because $1/s$ is an hyperbola, as in hyperbolic discounting. The authors do not make this connection but it is quite striking to see that they obtain the same result as Jones and Rachlin (2006) and Rachlin and Jones (2008), discussed in Section 2.2.

This approach has the advantage of using real friendships, hence increasing realism, but controlling for it with the network structure. In other words, existing social distance is used within the laboratory but objective measures are added to the picture through the network elicitation.

4.1.3 The spectrum of social distances

From this, it can be seen that there are mainly five ways to interpret social distance: anonymity, identification, common characteristics, minimal groups and networks. These can be conveniently organised by increasing realism and decreasing control. Varying anonymity to create social distances does not use any real-life, pre-existing social ties; hence it allows a very high level of control but it does not really map to real-world interactions, which is addressed by identification. On the other side of the spectrum, social distance as networks has the big advantage to use existing social ties. The other side of the coin is that studies using networks become vulnerable to usual criticisms: reputation and fear of punishment enter the picture, subjects are not playing the simple dictator game you show them but might fear the consequences etc.

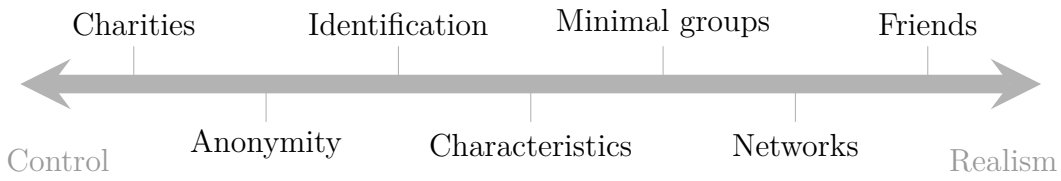


Figure 1: The spectrum of social distances.

We would like to add two extreme cases. The highest degree of control may be achieved if the questions do not involve other subjects at all. The questions might involve disembodied entities such as charities, organisations or government agencies. In the example of charities, the allocation questions might be of the form 'do you prefer to give $x^{\$}$ to charity X or x^{C} to charity Y?' The highest degree of realism can be achieved simply by asking groups of friends to come into the laboratory. This way, we use pre-existing social distances in the laboratory to ask allocation questions. This has also the advantage of capturing several dimensions of social distance.

To conclude, we came across several possible implementations of social distance. They all present advantages and disadvantages and can be organised in terms of a trade-off between control and realism. Figure 1 summarises the seven possible implementations of social distance that we have identified.

4.1.4 Social distance in the standard preference reversal experiment

Depending on which implementation of social distance we choose from Figure 1, the procedure will differ.

Even if it is an interesting research question, our experiment is not aimed at comparing different methods for having social distances in the laboratory. Hence, we pinned down a particular implementation and we decided to go for charities as a first attempt. The reason is the following. If we do not observe the social preference reversal using charities as social distance, our hunch is that it is unlikely that adding realism (moving to the right of Figure 1) would change anything. For example, it seems unlikely that we find no social preference reversal using charities but we do find a social preference reversal using groups of friends. So, for the time being, we take social distance as being the distance between a subject and a 'disembodied body', i.e. a charity.

4.2 Social preference reversal and charities: a pilot experiment

We carried a pilot experiment on the 29th of November 2013. It involved 18 subjects[2] divided in two sessions of 9 subjects. This was a pen-and-paper experiment. Each session

[2]This may sound like a too small sample. This is true and running more sessions is one of our priorities. In the meantime, as explained below there are 36 exploitable observations per subject, mitigating this caveat.

lasted roughly one hour. The payment was a fixed payment of £8. The instructions can be found in the Appendix. What follows is essentially an adaptation of what has been discussed in Section 3 to the specific case of using charities to create a social distance.

Essentially, we try to generate a choice/valuation preference reversal in the social domain. We use charities to create the social distance: there are 'close' charities, that is charities that are socially distant to the subject, and 'distant' charities. We use a tool developed in social psychology to check whether or not we were successful in creating different social distances with the charities we hand-picked. Subjects undertake two tasks: they choose between giving £5 to a close charity or £10 to a distant charity; and they value separately those two allocations. We use the ordinal payoff scheme as the incentive mechanism. The next section will describe the results, but as an advertisement we can say that we do find a social preference reversal, that is subjects tend to choose the C-allocation but value more the $-allocation.

4.2.1 Research question

The experiment is designed to address the first research question exposed on page 11: can we generate a choice/valuation preference reversal in the social domain? In other words, we mirror classical experiments carried in the context of choice under uncertainty and intertemporal choice by adapting them to the context of interpersonal choice. This is intended to be the first step into the research agenda described earlier.

4.2.2 C-allocations and $-allocations with 'close' and 'distant' charities

We need two social distances, small and large, in order to generate the C-allocations and the $-allocations. The social distance is defined as being between a subject and a charity. Hence, we chose charities organised into two groups: *'close' charities* which serve to construct the C-allocations, and *'distant' charities* for the $-allocations. The close charities are meant to be socially closer to the subject than the distant charities. The charities used are presented in Table 4.

As it can be seen from this table, the group of close charities is composed of well-known charities with very laudable goals. In contrast, distant charities have much more specific goals. We chose and organised the charities in those two groups based on introspection and data provided by the Charities Aid Foundation[3]. The main problem is that the social distance should be independent of the subject's preference or the subject's knowledge in order not to introduce a confounding factor. Hence, for the group of distant charities we tried not to use laughable or obviously comical charities (e.g. the Hastings Badger Protection Society), or too small-scale charities (e.g. The Borrowers' Toy Library Of Ripon). We screened small charities by their small number of volunteers or the low

[3] `http://www.charitytrends.org/`, accessed on 10 February 2014.

Table 4: List of charities (the abbreviations will be used in the remaining of the paper).

Close charities:
Save The Children (STC)
The Royal National Lifeboat Institution (RNLI)
Oxfam
The British Red Cross Society (BRCS)
The Salvation Army (SA)
The Royal Society For The Protection Of Birds (RSPB)
Distant charities:
The British Association for Shooting and Conservation (BASC)
Birdline UK Parrot Rescue (BUPR)
The Shooting Star Children's Hospice (SSCH)
The War Office Locomotive Trust (WOLT)
Red Squirrel Survival Trust (RSST)
London Wildlife Trust (LWT)

volume of donations they receive, but still with a real goal. Also, we invited only British students.

To test whether our selection has been successful or not, we relied on the *Inclusion of the Ingroup in the Self (IIS)* measure, presented in Figure 2. This measure was developed by Tropp and Wright (2001) as an extension of the Inclusion of the Other in the Self (IOS) measure by Aron et al. (1992), who introduced this measure in social psychology specifically as a tool to capture the different dimensions of social distance with a single number. They demonstrated its successfulness in doing so by showing that it cross-correlates with objective and subjective measures of closeness. Objective measures tackle the interactive and the habitual dimension of social distance: it can be said objectively that two individuals are in some ways socially close if they meet every day for some time and if they share cultural references. Subjective measures such as empathy are more affective-based. Tropp and Wright (2001) followed the same route and showed that the IIS measure has the same desirable properties as the IOS measure, making it a tool of choice for measuring the social distance between an individual and a group.

With 6 charities in each group, we have a total of 36 charity pairs with each pair being composed of one close charity and one distant charity. The C-allocations always provide $x^C = £5$ to a close charity while the \$-allocations provide $x^{\$} = £10$ to a distant charity[4]. To fix ideas, an example of a C-allocation is: 'give £5 to Save The Children'; and an example of a \$-allocation is 'give £10 to The War Office Locomotive Trust'.

[4]We choose multiples of 5 to be able to send money by post; see the incentive mechanism below.

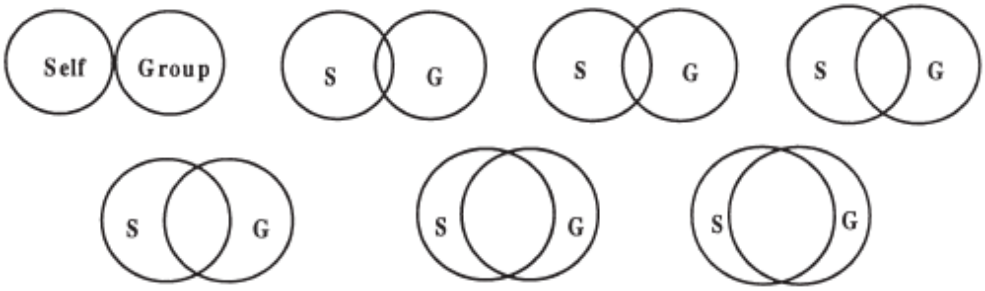


Figure 2: The IIS measure from Tropp and Wright (2001). "Respondents were asked to circle the pair of circles that 'you feel best represents your own level of identification with your group.' S=self, G=group" (note taken from the paper)

4.2.3 Choices, valuations and IIS measures

For each charity pair, we ask the subject to make a choice between the C-allocation and the $-allocation – this means a subject undertakes 36 choices.

Also, for each C-allocation and $-allocation, we elicit a valuation. To mirror the existing literature on choice/valuation preference reversals, we simply ask questions of the type: 'How much money given to you would be just as good as us giving £y to *charity X*?' $y = 5$ if charity X is a close charity and $y = 10$ if it is a distant charity. Thus, each subject provides 12 valuations.

Finally, we collect IIS measures, one for each charity. We ask subjects to draw a line around a pair of circle that they feel best represent how much they care for a given charity. This generates a one-to-seven scale that we take as our measure of social distance.

4.2.4 Tests

First of all, if our selection of close charities and distant charities has been accurate, IIS measures should be higher in the group of close charities than in the group of distant charities.

Then, for a given charity pair we say there is a *standard social preference reversal* – or social preference reversal for simplicity – if:

- $iis^C > iis^\$$, $(x^C, iis^C) \succ (x^\$, iis^\$)$ and $V\Big((x^\$, iis^\$)\Big) > V\Big((x^C, iis^C)\Big)$

In words: there is a preference reversal when a subject gives a higher IIS measure to a charity, chooses the allocation concerning this charity but value more the other allocation. We add the requirement on IIS measures to control for individual heterogeneity: it might happen that a subject, when presented a C-allocation and a $-allocation, feels that the charity in the $-allocation is socially closer.

This describes what we count as a social preference reversal and cannot be explained by standard economic theory. Critiques of this approach sometimes claim that this is a mere

experimental artefact caused by simple misunderstanding or mistakes on the subjects' side. To address this comment, we define *counter social preference reversal* as:

- $iis^C > iis^\$$, $(x^\$, iis^\$) \succ (x^C, iis^C)$ and $V\Big((x^C, iis^C)\Big) > V\Big((x^\$, iis^\$)\Big)$

If the explanation to social preference reversals were really mistakes or confusion, the proportions of social preference reversals and counter social preference reversals should be equal, because there is no reason to observe one more than the other. If, on the contrary, there were more social preference reversals than counter ones, then we have a stable and predictable pattern which cannot ne attributed to simple confusion and which is not addressed by standard economic theory.

4.2.5 Incentive mechanism: the ordinal payoff scheme

When one chooses an incentive mechanism, one wants it to be incentive compatible, that is to elicit truthful answers from the subjects. For this reason, we decided to implement the ordinal payoff scheme (Tversky et al., 1990; Cubitt et al., 2004). Under this scheme, subjects undertake a variety of tasks where they have to deal with a variety of objects. Let us take the example of classical preference reversals experiment to make the exposition clearer. In those experiments, subjects face lotteries (the objects) that they have to chose and to value (the tasks). They are told that, a lottery pair will be selected at the end of the experiment. Then, one of the two tasks (choice or valuation) will be selected and played for real. So, if the choice task is selected, the lottery they chose will be played; if the valuation task is selected, the lottery for which they gave the higher valuation will be played for real. Contrary to other mechanisms, e.g. the Becker-DeGroot-Marschak mechanism, the ordinal payoff scheme does not require the subjects to be expected utility maximisers – more specifically to satisfy the independence axiom – hence it does not introduce a further confounding factor. The only limitation of the ordinal payoff scheme is that valuations themselves are not incentivised, only the ranking they generate is. A subject understanding correctly the scheme should realise that valuations are only used to rank the alternatives but have no real meaning such as selling prices.

In our case, the ordinal payoff scheme implies that subjects are told a charity pair will be selected at the end of the experiment, and that their answer to the choice task or the valuation task will be randomly implemented. This triggers a second requirement: responses *have to be* implemented. For our experiment, this means that the money should actually be donated. This is a credibility requirement: if subjects think we are not actually going to give the money to the charity of their choice, then we cannot take their answers at face value. Thus, the instructions made very clear that the selected donation will be put in an envelope which they will close themselves. They also had to sign a receipt acknowledging the donation, and the delivery of the envelopes to the post-box where

monitored by a volunteer subject who also signed a receipt. This is also for this reason that we restricted donations to £5 and £10 to be able to send the money by post.

5 Results

This section describes the preliminary results from the pilot.

5.1 IIS measures

Table 5 and Figure 3 summarise the data on IIS measures. IIS measures organise the charities into two groups which correspond to our groups of close and distant charities. Thus, subjects' reported IIS measures validate our choice.

Interestingly, it seems that subjects did not hesitate to give the lowest IIS ratings to some charities: this can be seen by the fact that IIS ratings of distant charities are clustered around 1. This probably also explains why IIS ratings of distant charities have a lower dispersion, due to the censoring at 1. They were apparently reluctant to give too high IIS ratings: close charities are clustered around 3 and 4, and there is a higher dispersion for those charities.

There are two outliers: The Royal Society For The Protection Of Birds tended to be considered as a distant charity by the subjects; on the contrary, The Shooting Stars Children's Hospice was perceived as a close charity. However, this is not much of a problem since we control for reported IIS measures to compute the social preference reversals.

5.2 Are there social preference reversals?

This is indeed the main question we are interested in.

In total we have 18 (sujects) $\times$ 36 (charity pairs) = 648 observations. Table 6 reports frequencies of the four possible patterns of preference *for the whole sample* – hence without taking into account subjects' reported IIS measures. Consistent patterns are on the diagonal and (standard and counter) reversals are on the off-diagonal. Figure 4 reports the same information in a graphical way. Consistent patterns account for roughly 50% of all observations while social preference reversals constitute almost 43%. Also, counter preference reversals are definitely a minority: around 7%.

The conclusions do not change much if we restrict the analysis to the subsample where the reported IIS measure for the close charity is greater than the reported IIS measure for the distant charity. This is presented in Table 7 and Figure 5. Interestingly, the number of consistent patterns for the $-allocation drop between the whole sample and the subsample. This means that many subjects chose the $-allocation in the choice task despite giving a lower IIS measure to the distant charity (which is in the $-allocation) than to the close charity. The same conclusion holds if we restrict the analysis to the

Table 5: Summary statistics of IIS measures.

Charities	Mean	Std. Err.	95% Conf.	Interval
STC	4.833333	.3898215	4.010882	5.655785
RNLI	3.666667	.3429972	2.943006	4.390327
Oxfam	4	.4644204	3.020159	4.979841
BRCS	4.388889	.2930445	3.770619	5.007159
SA	3.833333	.3157106	3.167242	4.499425
RSPB	2.444444	.3544509	1.696618	3.19227
BACS	1.722222	.2258688	1.245681	2.198764
BUPR	2	.1980295	1.582194	2.417806
SSCH	4.5	.4731345	3.501774	5.498226
WOLT	1.944444	.2855132	1.342064	2.546825
RSST	2.222222	.190554	1.820188	2.624256
LWT	2.944444	.2209933	2.478189	3.4107

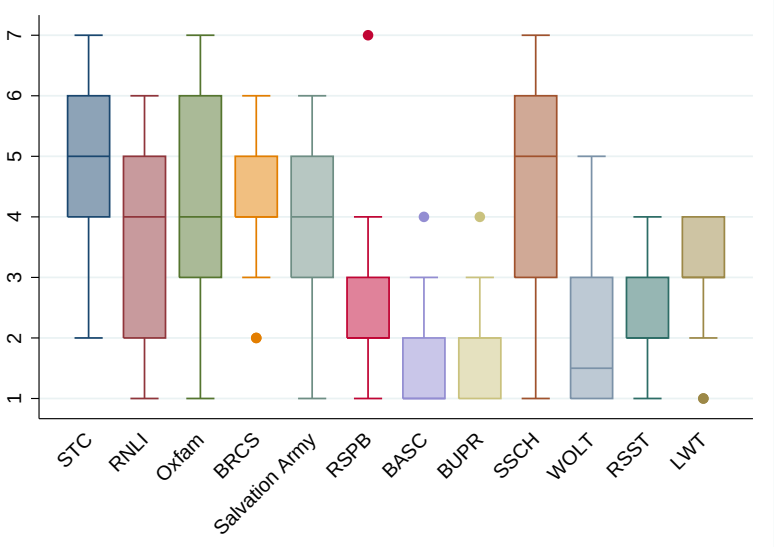


Figure 3: Boxplot of IIS measures.

Table 6: Patterns of preference, frequencies, whole sample.

Choice	*Valuation*		
	Value more $	Value more C	Total
Choose $	141	47	188
Choose C	278	182	460
Total	419	229	648

subsample where there is at least a difference of 2 between the IIS measure of the close charity and the IIS measure of the distant charity (not displayed).

t-tests reject the null hypothesis that preferences revealed in the choice task and in the valuation task are the same, both in the whole sample and in the subsample ($t = 14.82$ and 20.57 with $p < 0.001$, respectively). The conclusion does not change when one clusters at the level of the subject to account for correlation, even if t-values are much lower ($t = 5.66$ and 5.58 with $p < 0.001$, respectively). However, it should be noted that this aggregate testing hides disparities between the charity pairs. Table 8 reports t-tests carried independently for each charity pair. While H_0 can be rejected at the 0.05 level for most charity pairs, some charities seem to counteract the preference reversal effect, notably SSCH. This is confirmed by the previous analysis of IIS measures: SSCH tended to be perceived as a close charity by the subjects.

5.3 Next steps

We observe social preference reversals in our pilot experiment, and despite the obvious need of running more sessions it is very unlikely that this clear pattern will fade away when more data will be collected. However, some questions remain unanswered, and especially: why do we obtain such a social preference reversal?

The literature has identified that preference reversals are mainly caused by valuation tasks (Tversky et al., 1990). Introspection tells us that it might be even the more true in our case. The reason is that valuation tasks have a direct interpretation in the context of choice under uncertainty as selling prices, but this natural interpretation is lacking in here. What is really an equivalence valuation of an allocation? The question puzzles us, and it would not be surprising if it puzzled also the subjects, potentially leading to unreliable results.

Thus, our strategy for future sessions is to try to reduce subjects' cognitive strain by turning off some features of our design to make it more understandable. There are two elements that we think can add misunderstanding to the picture: the ordinal payoff scheme and the fact that valuation tasks mix money for a charity and money for the subject. Hence we will run two treatments: one with purely hypothetical questions (money will not be donated to the charities), and one using a reference charity and not the subject in the valuation task. We hope tu run those treatments in the next weeks.

6 Conclusion

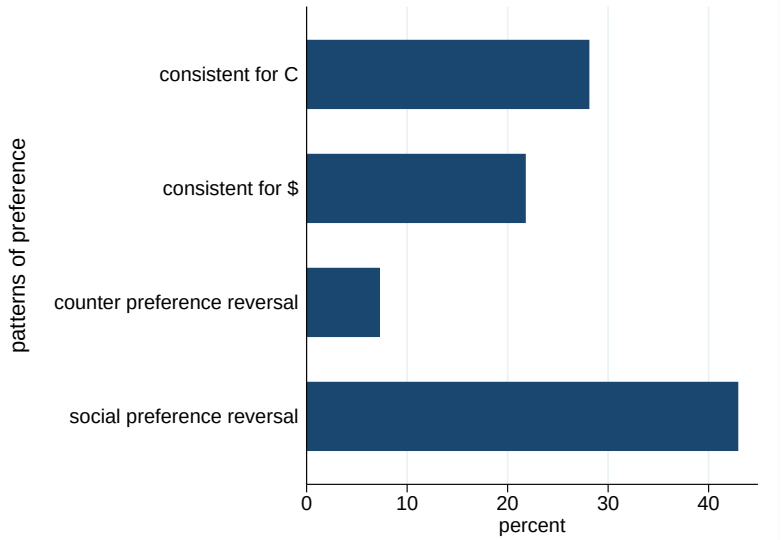


Figure 4: Patterns of preference, whole sample.

Table 7: Patterns of preference, frequencies, subsample.

Choice	*Valuation*		
	Value more $	Value more C	Total
Choose $	34	8	42
Choose C	237	153	390
Total	271	161	432

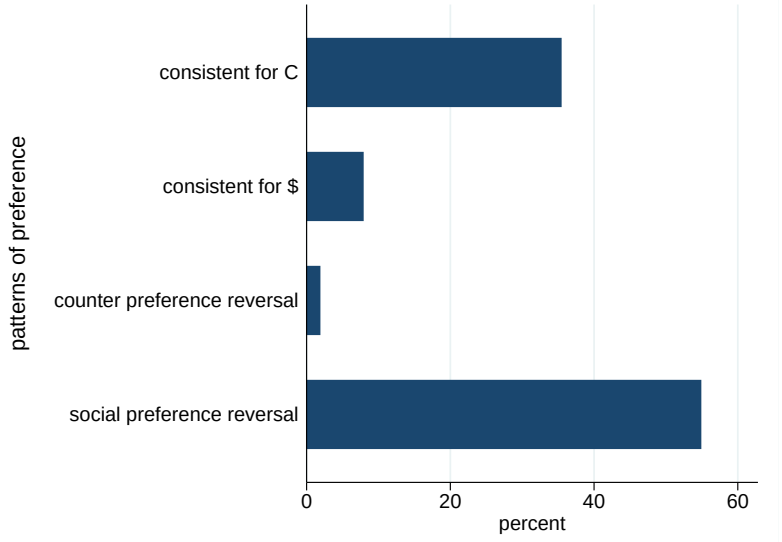


Figure 5: Patterns of preference, subsample.

Table 8: t-tests for each charity pair, H_0: preference revealed in the choice task and in the valuation task are the same.

Charity pair	t	p
STC - BASC	5.83	<0.001
STC - BUPR	5.17	<0.001
STC - SSCH	0	1
STC - WOLT	4.27	<0.001
STC - RSST	4.61	<0.001
STC - LWT	5.17	<0.001
RNLI - BASC	3.06	0.007
RNLI - BUPR	4.61	<0.001
RNLI - SSCH	-0.80	0.430
RNLI - WOLT	3	0.008
RNLI - RSST	3.06	0.007
RNLI - LWT	2.71	0.015
Oxfam - BASC	5.83	<0.001
Oxfam - BUPR	4.61	<0.001
Oxfam - SSCH	-2.20	0.042
Oxfam - WOLT	3.83	0.001
Oxfam - RSST	3.69	0.002
Oxfam - LWT	3.69	0.002
BRCS - BASC	5.17	<0.001
BRCS - BUPR	4.61	<0.001
BRCS - SSCH	-0.80	0.430
BRCS - WOLT	4.61	<0.001
BRCS - RSST	3.06	0.007
BRCS - LWT	3.69	0.002
SAL - BASC	5.83	<0.001
SAL - BUPR	4.12	<0.001
SAL - SSCH	-2.20	0.042
SAL - WOLT	3.43	0.003
SAL - RSST	4.61	<0.001
SAL - LWT	3.29	0.004
RSPB - BASC	1.14	0.269
RSPB - BUPR	2.38	0.029
RSPB - SSCH	-3.43	0.003
RSPB - WOLT	1.14	0.269
RSPB - RSST	0.70	0.496
RSPB - LWT	-0.57	0.579

References

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K Wagner (2011), "Social preferences and self-control." Working paper, SSRN.

Albert, Stuart (1977), "Temporal comparison theory." *Psychological Review*, 84, 485–503.

Aron, Arthur, Elaine N Aron, and Danny Smollan (1992), "Inclusion of other in the self scale and the structure of interpersonal closeness." *Journal of Personality and Social Psychology*, 63, 596–612.

Baucells, Manel and Franz H Heukamp (2010), "Common ratio using delay." *Theory and Decision*, 68, 149–158.

Baucells, Manel and Franz H Heukamp (2012), "Probability and time trade-off." *Management Science*, 58, 831–842.

Baumeister, Roy F, Ellen Bratslavsky, Mark Muraven, and Dianne M Tice (1998), "Ego depletion: Is the active self a limited resource?" *Journal of Personality and Social Psychology*, 74, 1252–1265.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1964), "Measuring utility by a single-response sequential method." *Behavioral Science*, 9, 226–232.

Benzion, Uri, Amnon Rapoport, and Joseph Yagil (1989), "Discount rates inferred from decisions: An experimental study." *Management Science*, 35, 270–284.

Bogardus, Emory S (1925), "Measuring social distance." *Journal of Applied Sociology*, 9, 299–308.

Bohnet, Iris and Bruno S Frey (1999a), "Social distance and other-regarding behavior in dictator games: Comment." *American Economic Review*, 89, 335–339.

Bohnet, Iris and Bruno S Frey (1999b), "The sound of silence in prisoner's dilemma and dictator games." *Journal of Economic Behavior & Organization*, 38, 43–57.

Bolton, Gary E, Elena Katok, and Rami Zwick (1998), "Dictator game giving: Rules of fairness versus acts of kindness." *International Journal of Game Theory*, 27, 269–299.

Chapman, Gretchen B and Bethany J Weber (2006), "Decision biases in intertemporal choice and choice under uncertainty: testing a common account." *Memory & Cognition*, 34, 589–602.

Charness, Gary and Uri Gneezy (2008), "What's in a name? anonymity and social distance in dictator and ultimatum games." *Journal of Economic Behavior & Organization*, 68, 29–35.

Charness, Gary, Ernan Haruvy, and Doron Sonsino (2003), "Social distance and reciprocity: The internet vs. the laboratory." Working paper, University of California Santa Barbara.

Charness, Gary, Luca Rigotti, and Aldo Rustichini (2007), "Individual behavior and group membership." *American Economic Review*, 97, 1340–1352.

Chen, Yan and Sherry Xin Li (2009), "Group identity and social preferences." *American Economic Review*, 99, 431–57.

Cubitt, Robin P, Alistair Munro, and Chris Starmer (2004), "Testing explanations of preference reversal." *Economic Journal*, 114, 709–726.

Cubitt, Robin P, Chris Starmer, and Robert Sugden (1998), "Dynamic choice and the common ratio effect: An experimental investigation." *Economic Journal*, 108, 1362–1380.

Curry, Oliver S, Michael E Price, and Jade G Price (2008), "Patience is a virtue: Cooperative people have lower discount rates." *Personality and Individual Differences*, 44, 780–785.

Dufwenberg, Martin and Astri Muren (2006), "Generosity, anonymity, gender." *Journal of Economic Behavior & Organization*, 61, 42–49.

Eckel, Catherine C and Philip J Grossman (2005), "Managing diversity by creating team identity." *Journal of Economic Behavior & Organization*, 58, 371–392.

Epper, Thomas, Helga Fehr-Duda, and Adrian Bruhin (2011), "Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting." *Journal of Risk and Uncertainty*, 43, 169–203.

Etang, Alvin, David Fielding, and Stephen Knowles (2011), "Does trust extend beyond the village? experimental trust and social distance in cameroon." *Experimental Economics*, 14, 15–35.

Festinger, Leon (1954), "A theory of social comparison processes." *Human Relations*, 7, 117–140.

Fishburn, Peter C and Ariel Rubinstein (1982), "Time preference." *International Economic Review*, 23, 677–694.

Frohlich, Norman, Joe Oppenheimer, and J Bernard Moore (2001), "Some doubts about measuring self-interest using dictator experiments: The costs of anonymity." *Journal of Economic Behavior & Organization*, 46, 271–290.

Frohlich, Norman, Joe Oppenheimer, and Anja Kurki (2004), "Modeling other-regarding preferences and an experimental test." *Public Choice*, 119, 91–117.

Glaeser, Edward L, David I Laibson, Jose A Scheinkman, and Christine L Soutter (2000), "Measuring trust." *Quarterly Journal of Economics*, 115, 811–846.

Goeree, Jacob K, Margaret A McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv (2010), "The 1/d law of giving." *American Economic Journal: Microeconomics*, 2, 183–203.

Green, Leonard, Nathanael Fristoe, and Joel Myerson (1994), "Temporal discounting and preference reversals in choice between delayed outcomes." *Psychonomic Bulletin & Review*, 1, 383–389.

Grether, David M and Charles R Plott (1979), "Economic theory of choice and the preference reversal phenomenon." *American Economic Review*, 69, 623–638.

Halevy, Yoram (2008), "Strotz meets allais: Diminishing impatience and the certainty effect." *American Economic Review*, 98, 1145–1162.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L Smith (1994), "Preferences, property rights, and anonymity in bargaining games." *Games and Economic Behavior*, 7, 346–380.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L Smith (1996), "Social distance and other-regarding behavior in dictator games." *American Economic Review*, 86, 653–660.

Johannesson, Magnus and Björn Persson (2000), "Non-reciprocal altruism in dictator games." *Economics Letters*, 69, 137–142.

Jones, Bryan and Howard Rachlin (2006), "Social discounting." *Psychological Science*, 17, 283–286.

Jones, Bryan and Howard Rachlin (2009), "Delay, probability, and social discounting in a public goods game." *Journal of the Experimental Analysis of Behavior*, 91, 61–73.

Kahneman, Daniel and Amos Tversky (1979), "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–291.

Kanfer, Frederick H (1979), "Personal control, social control, and altruism: Can society survive the age of individualism?" *American Psychologist*, 34, 231–239.

Karakayali, Nedim (2009), "Social distance and affective orientations." *Sociological Forum*, 24, 538–562.

Keren, Gideon and Peter Roelofsma (1995), "Immediacy and certainty in intertemporal choice." *Organizational Behavior and Human Decision Processes*, 63, 287–297.

Kirby, Kris N and Richard J Herrnstein (1995), "Preference reversals due to myopic discounting of delayed reward." *Psychological Science*, 6, 83–89.

Kirby, Kris N and Nino N Maraković (1995), "Modeling myopic decisions: Evidence for hyperbolic delay-discounting within subjects and amounts." *Organizational Behavior and Human Decision Processes*, 64, 22–30.

Kocher, Martin, Kristian Ove R Myrseth, Peter Martinsson, and Conny E Wollbrant (2012), "Strong, bold, and kind: Self-control and cooperation in social dilemmas." Working paper no 523, University of Gothenburg.

Lichtenstein, Sarah and Paul Slovic (1971), "Reversals of preference between bids and choices in gambling decisions." *Journal of Experimental Psychology*, 89, 46–55.

Lichtenstein, Sarah and Paul Slovic (1973), "Response-induced reversals of preference in gambling: An extended replication in las vegas." *Journal of Experimental Psychology*, 101, 16–20.

Lindman, Harold R (1971), "Inconsistent preferences among gambles." *Journal of Experimental Psychology*, 89, 390–397.

Liviatan, Ido, Yaacov Trope, and Nira Liberman (2008), "Interpersonal similarity as a social distance dimension: Implications for perception of others' actions." *Journal of Experimental Social Psychology*, 44, 1256–1269.

Martinsson, Peter, Kristian Ove R Myrseth, and Conny E Wollbrant (2010a), "Conditional cooperation: Evidence for the role of self-control." Working paper no 10-004, European School of Management and Technology.

Martinsson, Peter, Kristian Ove R Myrseth, and Conny E Wollbrant (2010b), "Reconciling pro-social vs. selfish behavior: Evidence for the role of self-control." *Judgment and Decision Making*, 7, 304–315.

Meer, Jonathan and Oren Rigbi (2011), "Transactions costs and social distance in philanthropy: Evidence from a field experiment." Discussion paper no. 12-05, Monaster Center for Economic Research.

Myrseth, Kristian Ove R and Ayelet Fishbach (2009), "Self-control a function of knowing when and how to exercise restraint." *Current Directions in Psychological Science*, 18, 247–252.

Prelec, Drazen and George Loewenstein (1991), "Decision making over time and under uncertainty: A common approach." *Management Science*, 37, 770–786.

Pronin, Emily, Christopher Y Olivola, and Kathleen A Kennedy (2008), "Doing unto future selves as you would do unto others: Psychological distance and decision making." *Personality and Social Psychology Bulletin*, 34, 224–236.

Pronin, Emily and Lee Ross (2006), "Temporal differences in trait self-ascription: when the self is seen as an other." *Journal of Personality and Social Psychology*, 90, 197–209.

Quiggin, John and John Horowitz (1995), "Time and risk." *Journal of Risk and uncertainty*, 10, 37–55.

Rachlin, Howard and Leonard Green (1972), "Commitment, choice and self-control." *Journal of the Experimental Analysis of Behavior*, 17, 15–22.

Rachlin, Howard and Bryan Jones (2008), "Social discounting and delay discounting." *Journal of Behavioral Decision Making*, 21, 29–43.

Rachlin, Howard, Alexandra W Logue, John Gibbon, and Marvin Frankel (1986), "Cognition and behavior in studies of choice." *Psychological Review*, 93, 33–45.

Rachlin, Howard, Andres Raineri, and David Cross (1991), "Subjective probability and delay." *Journal of the Experimental Analysis of Behavior*, 55, 233–244.

Rachlin, Howard and Eric Siegel (1994), "Temporal patterning in probabilistic choice." *Organizational Behavior and Human Decision Processes*, 59, 161–176.

Slovic, Paul, Dale Griffin, and Amos Tversky (1990), "Compatibility effects in judgment and choice." In *Insights in Decision Making: Theory and Application* (R. M. Hogarth, ed.), 5–27, The University of Chicago Press.

Starmer, Chris and Robert Sugden (1989), "Violations of the independence axion in common ratio problems: An experimental test of some competing hypotheses." *Annals of Operations Research*, 19, 79–102.

Tajfel, Henri, Michael G Billig, Robert P Bundy, and Claude Flament (1971), "Social categorization and intergroup behaviour." *European Journal of Social Psychology*, 1, 149–178.

Tajfel, Henri and John C Turner (1979), "An integrative theory of intergroup conflict." In *The Social Psychology of Intergroup Relations* (W. G. Austin and S. Worchel, eds.), 47, Brooks/Cole Publishing Company.

Thaler, Richard H (1981), "Some empirical evidence on dynamic inconsistency." *Economics Letters*, 8, 201–207.

Todorov, Alexander, Amir Goren, and Yaacov Trope (2007), "Probability as a psychological distance: Construal and preferences." *Journal of Experimental Social Psychology*, 43, 473–482.

Trope, Yaacov and Nira Liberman (2010), "Construal-level theory of psychological distance." *Psychological Review*, 117, 440–463.

Tropp, Linda R and Stephen C Wright (2001), "Ingroup identification as the inclusion of ingroup in the self." *Personality and Social Psychology Bulletin*, 27, 585–600.

Tversky, Amos, Paul Slovic, and Daniel Kahneman (1990), "The causes of preference reversal." *American Economic Review*, 80, 204–17.

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

A Instructions

Welcome to the experiment. In this experiment, we are going to ask you questions about charities. You will have the opportunity to make a donation to a charity of your choice. To be more precise, depending on your choices, we are going to give money to charities on your behalf. Hence, *the money donated to the charities is not your money.* You will also receive a fixed payment of £8. The experiment will last less than one hour.

Questions

There will be three types of tasks, which we call **donation tasks**, **equivalence tasks** and **circle tasks**. We are going to describe each of them in more detail.

Donation tasks

Donation tasks ask you to choose between two alternatives. Here is an example:

> **Option A**: We give £5 to *charity X*
> **Option B**: We give £10 to *charity Y*
>
> Choose **A** or **B**:

You will choose one of the two options by writing 'A' or 'B'. There will be a range of such tasks for a variety of charities.

Equivalence tasks

In equivalence tasks, we propose a donation of a specific amount (say, £y) to a specific charity (say, charity X). We then ask you to specify how much money we would have to give to you so that you would think that amount of money was just as good as the proposed donation. Here is an example of those questions:

> How much money given to you would be just as good as us giving £5 to *charity X*?
>
> Please write the amount here:
>
> We will refer to this amount as your equivalence valuation of giving £5 to charity X.

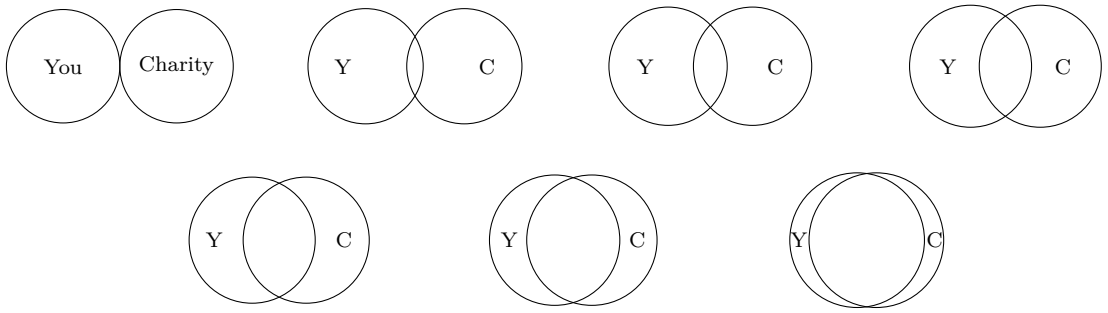
Charity X and the amount of the donation will change from task to task.

We will never give money to you as part of this task, but as explained later your answer may affect which charity we are actually going to give money to, so please give considered and careful answers.

Circle tasks

For those tasks, we will ask you how much you care for a charity. You will be asked in this way:

Please consider charity X. Select the pair of circles that you feel best represents how much you care for *charity X* (Y=You, C=charity X):



Indicate your answer by drawing a line around the pair of circles you select.

Procedure

On your left is a pile of paper containing several of the aforementioned tasks in a random order. Once instructed by the experimenter, you may take the first strip of paper. When you have completed the task on it, please enclose it in the envelope placed on your right and do not consider it again. You may then take another strip of paper, answer it, place it in the envelope, and repeat until you have exhausted the pile on your left. Once you have exhausted the pile of paper, please raise your hand. You will not be allowed to leave the room until everyone has finished.

The charity to which we are going to give the money on your behalf will be selected as follows. At the end of the experiment, the experimenter will come to your table and will randomly select two charities W and Z by drawing them out of a bag. You would have encountered each of those two charities twice:

- both of them at the same time in a *donation task*, where you had to choose between giving to charity W or to charity Z;
- each of them separately in *equivalence tasks*, where you had to tell us the amount of money that makes you equally happy between giving to charity W or Z and getting money for you.

We will then flip a coin. If the coin comes up heads, we will give the money to the charity you selected in the donation task. If it comes up tails, we will give the money to the charity you gave a higher equivalence valuation to.

Let us illustrate this with an example. Imagine two fictious charities: Nolfolck and UCAR. When the experimenter comes to your table at the end of the experiment, imagine that this pair of charities is randomly drawn from the bag. During the experiment, you encountered Nolfolck and UCAR in the following donation task:

Option A: We give £5 to Nolfolck
Option B: We give £10 to UCAR

Choose **A** or **B**: A

Suppose you decided to choose **A**, that is, giving £5 to Nolfolck.

You also faced an equivalence task with Nolfolck:

How much money given to you would be just as good as us giving £5 to Nolfolck?

Please write the amount here: 5

We will refer to this amount as your equivalence valuation of giving £5 to Nolfolck.

This example supposed that you stated that Norfolck having £5 and you getting £5 makes you equally happy. You faced a similar question with UCAR:

How much money given to you would be just as good as us giving £10 to UCAR?

Please write the amount here: 7

We will refer to this amount as your equivalence valuation of giving £10 to UCAR.

Here, imagine that you stated that UCAR having £10 and you getting £7 makes you equally happy.

Finally, we use a coin flip to determine whether it is your response to the choice task, or your responses to the equivalence tasks that determine the charity we will pay on your behalf:

- If the donation task is selected, we are going to give £5 to Nolfolck on your behalf because this is what you chose in the donation task involving Nolfolck and UCAR.
- If the equivalence task is selected, we are going to give £10 to UCAR because your equivalence valuation of giving £10 to UCAR (£7) is greater than your equivalence valuation of giving £5 to Nolfolck (£5).

At this stage, we will know the charity selected and the amount donated. We will place your donation in a stamped envelope with the address of the selected charity and we will ask you to close the envelope. You will receive your show-up fee, and we will ask you to sign a receipt acknowledging that you received it and that you gave a certain amount of money to the selected charity.

Finally, we will collect all the envelopes and, with a volunteer, we will go deposit them in the postbox situated in Portland Building.

* * *

The experiment will begin soon. If you have any question, please raise your hand and the experimenter will come to your table to answer it. Please do not look at the questions before instructed to do so.