ESSAYS IN BEHAVIOURAL, ECONOMIC, AND COGNITIVE SCIENCES

By

Eleni Achtypi

A thesis submitted in partial fulfilment of

the requirements for the degree of Doctor of Philosophy

in Psychology

University of Warwick, Department of Psychology

October 2018

Table of Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I would like to thank my supervisor Gordon D.A. Brown for his insight, ideas, guidance, direction and continuous support. I constantly feel extremely inspired by the way he thinks about research, his intuition, and how he creates scientific connections. He has been the biggest influence in my academic studies and I will be forever grateful. It has been an honor and pleasure working with him. I would like to extend my appreciation to my second supervisor, Lukasz Walasek who has also been extremely supportive and caring, and his enthusiasm about Behavioural Science has been infectious. I am also very grateful to Alex Mushore and Ty Hayes who helped me with recruiting and running some of my experiments at the Warwick Business School Behavioural Science Laboratories and to the Leverhulme Trust which has been instrumental in providing me the funding to carry out research in Behavioural, Economic, and Cognitive Sciences.

The Psychology Department and its administration has been extremely helpful to me throughout my research. I would particularly like to thank Nicky, Steve and Kaushalya for helping me deal with administrative and financial issues involved in running my experiments and for always having a smile on their faces. I would also like to thank Joe from MTurk Data who made sure that some of my online studies reached the participants smoothly. Thank you to all the friends and PhD students from diverse research backgrounds with whom I have had great research discussions. Special thanks go to Edika, Ayse and Neel. I would also like to thank all the supportive and caring people I have met in Philadelphia during the last months of writing up my thesis. This space is too small to list them all. I feel extremely lucky to have such wonderful friends and family. The Vlachou family; Mimis, Vivi, Yannis and Magdalina are like a second family to me and I am deeply grateful for their support throughout my research. I want to also thank my godmother Eleni, Elpida, Popi and Evi for being incredibly loving and supportive friends to me and my family.

Lastly, I would like to thank my wonderful family. My precious mother Anthi and my beloved grandmother Eleni for being my rocks throughout this process, and my sister Katerina for teaching me how to be brave. My partner Tom, thank you for your immense love, support and care, I am extremely grateful that you and my family have always believed in me and been by my side.

This is dedicated to my γιαγιά (granny)
Eleni who is no longer with us.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Psychology. I hereby declare that the research reported in this thesis is my own work unless otherwise stated. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author, except for Table 1.4, results of which were produced by Gordon D.A. Brown.

Part of Chapter 1 has been prepared as a manuscript in collaboration with Simon Gächter and Gordon D.A. Brown.

A Rank-based Model of Social Conformity in the Voluntary Provision of Public Goods, Achtypi E., Brown G.D.A., & Gächter S.

Part of Chapter 2 has been prepared as a manuscript and submitted to the *Journal of Consumer Psychology* in collaboration with Lukasz Walasek, Nathaniel J. S. Ashby, Eldad Yechiam, and Gordon D.A. Brown.

The endowment effect reflects concern with good dealness, not changes in underlying valuations, Achtypi E., Walasek L., Ashby N. J. S., Brown G. D. A., & Yechiam E.

Chapter 3 has been prepared as a working paper with Lukasz Walasek and Gordon D.A. Brown.

Perceptions of Economic Inequality and Subjective Well-Being, Achtypi E., Walasek L., & Brown G. D. A.

Abstract

This thesis is comprised of three separate chapters in Behavioural Science. The general underlying theme of the seven studies presented in this thesis involves a rank approach and influence in the elicitation and response to contextual distributions. Specifically, Chapter 1 addresses the influence of other people's contributions (created to represent a diverse range of distributions of others' contribution decisions) on aggregate cooperation and individual contributions in public goods games. It was found that participants take into account and are influenced by subtle aspects of what others contribute, providing a basis for the model we developed according to which participants have a preference for placing themselves at a specific relative ranked position within their inferred social norm of others' contributions. Chapter 2 discusses the elicitation of participants' market price distributions, quality estimates and market price estimates of a specific item in an endowment effect study in order to develop a model based on the quantification of "good dealness". The model assumes no cognitive bias or ownership-induced changes in underlying preferences. It was found that sellers demanded a market-appropriate price for the item given their beliefs about the item's relative quality and their beliefs about the distribution of market prices of similar items in the market. Buyers, in contrast, offered less than what they believed the appropriate market price to be because they would only offer a price that represented a good deal. Chapter 3 examines the elicitation of individuals' subjective estimates of entire income and wealth distributions and also tested the idea that the unclear pattern of findings between income inequality and individual subjective well-being in previous literature reflects individual differences in people's perception of income inequality. We found that on aggregate people are relatively accurate (and consistent across time) in their estimates of income and wealth inequality, although they overestimate incomes at all levels and overestimate low and high levels of wealth. However, there were some small individual differences in perceived inequality. Higher perceived inequality was associated with lower income/wealth and more liberal ideology. Various measures of subjective well-being were not predicted by perceived inequality of either income or wealth. Yet, people's subjective well-being was best predicted by their subjective estimates of where they rank in the population, with the overall pattern of findings being consistent with a model in which people care about their perceived relative rank position within income and wealth distributions.

Chapter 1:

Rank Effects and Context Effects within Social Norms in Public Goods Games

"You can discover more about a person in an hour of play than a year of conversation."-Plato.

1.1 Introduction

Most studies investigating behaviour in social dilemmas use four-player repeated public goods games in which participants only receive information about the average contributions of other players in the previous rounds. These are situations that produce conflict between what is rational for an individual and what is best for the group. Many of these studies find a positive relationship between an individual's contribution and the contributions made by others (Bardsley, 2000; Keser and van Winden, 2000; Weimann, 1994). On the basis of such experiments it is however not possible to determine whether participants are concerned with how their own contribution relates to the mean of others' contributions or with how their contribution ranks within an inferred distribution of others' contributions (the "social norm"). However, intuition, and experimental evidence in psychology, suggests that individuals may show rank-sensitivity in their contributions.

The aim of this chapter is therefore to examine the effects of context and social norms in the context of the frequently used public goods game, an abstract social dilemma decision task whereby participants are asked to contribute to a group pot. We intend to investigate how aspects of the other group members' behaviour affect one's decisions. Our goal is to examine more subtle effects of the context besides what is typically been studied (the mean of others' contribution behaviours). Moreover, we hypothesised that when participants see the whole distribution of others' contributions, they infer a social norm in which they have a preference to place themselves at a specific rank position.

The chapter is structured as follows. We first discuss some key definitions and the literature regarding the provision of information on public goods games. Second, we outline the details of the first study (P-experiment) we conducted, report the results, primarily on an aggregate level, and then formulate our rank-based model of decision making and demonstrate how it accounts for our observations (additional analysis regarding individual differences in the P-experiment is reported in the Appendix to this chapter). We then outline the details of our second study (C-experiment), report the results and conclude with a discussion.

1.1.1 Social Norms

People's choices, and the attitudes they express, are strongly influenced by social context — it has long been known that people compare themselves with others and often adjust their behaviour in the direction of a social norm (e.g., Buunk & Gibbons, 2007; Festinger, 1954; Hyman, 1942). More recently, much attention has been given to the idea that providing people with information about social norms can influence behaviour such as energy usage (Allcott & Rogers, 2014; Ayres, Raseman, & Shih, 2013; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007) and alcohol consumption (Neighbors, Larimer, & Lewis, 2004) with the mean energy consumption or alcohol consumption being provided as the social norm. Behaviour change achieved by giving people information about social norms is one category of "nudge" (Thaler & Sunstein, 2008) which may provide policy-makers with cost-effective ways of changing people's behaviour without restricting their freedom of choice (Benartzi et al., 2017).

Social norm effects are found in a wide variety of domains. Martin and Randal (2010) found that the size of visitors' voluntary donations to an art gallery were strongly and systematically influenced by the amount and typical denomination of previous donations (visible in a transparent donation box), and preferences for music are strongly influenced by social norms (Salganik, Dodds, & Watts, 2006). Chen, Harper, Konstan, and Li (2010) found that the number of ratings provided by the median user of a movie rating site could be increased by more than 500% by providing users with information about the number of ratings given by other people. Hallsworth et al. (2016) used suitably-framed messages based on social norms ("The great majority (80%) of practices in [NHS Area Team] prescribe fewer antibiotics per head than yours") as part of an intervention to reduce undesirable over-prescription of antibiotics. Local debt-norm messages that HMRC used in letters sent to customers led to increases in tax revenue of 210 million pounds, showcasing how social influence and subtle changes in the context of choice can affect behavior (BIT staff, 2013).

There is thus a wide variety of evidence that people's behaviour can be influenced by social norms, and such findings have important policy implications. There remains, however, a striking lack of theory of the underpinnings of social norm effects. In this chapter, we offer some initial steps towards a theoretical framework along with experimental evidence. We do so in the specific context of multi-person social dilemmas, where it is already known that provision of social norm information can influence cooperation and contributions (e.g., Fischbacher & Gächter, 2010). Our focus here is on descriptive social norms — information about other people's behaviour — although we note that the term "social norms" is used in a number of different ways in both the economic and psychological literatures (Bicchieri, 2006).

1.1.2 Cooperation and Social Dilemmas

Cooperation between individuals, a basic component of human society, typically occurs within a social context where other agents make similar decisions. Your decision of how much to donate to charity, for example, is commonly made with the knowledge, or at least perception, of how much others gave. For an individual the decision to cooperate, or not, is consequently based on their own internal preferences along with the influence of information they receive from the collective. Many studies of cooperative behaviour involve multi-person social dilemmas (Dawes, 1980), situations that raise a conflict between what is individually and collectively optimal (there are dilemmas involving only two people, such as the famous prisoner's dilemma: Rapoport & Chammah, 1965). If the most rational behaviour in economic terms is chosen by most individuals, then the collective suffers, and the opposite is also true. Abstract decision tasks inspired by game theory, coined "experimental games", have been used to model these situations either in the laboratory or the field. Participants typically have to choose between a range of cooperative and no cooperative strategies that will have consequences for them through their experimental payoffs (typically money or other prizes).

1.1.3 Types of multi-person Social Dilemma

The Public Goods dilemma and the Common Goods dilemma are common types of social dilemma that have been modelled as experimental games. Common goods possess two characteristics: They are rivalrous and non-excludable (Olson 1965; Samuelson, 1954). Consumption by one does prevent the consumption by another person but nobody can be excluded from consumption. The classic example is fish stocks. Nobody is excluded from fishing, but if there are no limits or regulations involved, the amount one person fishes now will influence the amount a future fisherman could fish. The "tragedy of the commons" (Hardin, 1968) occurs when we gain short-term benefits by extracting or overusing resources without acknowledging the long-term consequences. A public good on the other hand is a good that is both non-rivalrous and non-excludable. Everyone can consume it with no exclusions and one person's consumption does not affect or prevents another's. Nevertheless, if everybody free rides by not contributing to the good then the good will not be produced and no one will receive its benefits. National defense, street lighting, education and fire brigade are some examples of public goods. Here, we focus on public goods social dilemmas (although see section 1.6 for pilot studies that utilised a common pool resource dilemma).

1.1.4 Public Goods Games-Decision Situation

Public goods games can be thought as creating positive externalities. In such games typically, the linear voluntary contribution mechanism is used. Participants are split into groups of typically $\mathrm{N} = 4$ people. Each member receives an endowment typically of $\mathrm{e} = 20$ points and must decide how many of these points to allocate to their private account, which yields constant returns and how many to contribute to the public account which yields benefits to all group members. Each member's payoff is then determined by

$$\pi_i = e - x_i + \tau \sum_{j=1}^{N} x_j, \qquad (1)$$

where x_i is the amount player *i* contributes to the public good. The marginal per capita rate of return, τ, is typically set at 0.5 and cannot exceed 1 or be less than $1/\mathrm{N}$. This is because having $\tau < 1$ sets the dominant strategy to be not contributing at all to the public good, while $\tau > 1/\mathrm{N}$ will maximise the aggregate payoff if each player contributes e. With this kind of parameterisation the Nash (individually optimal) equilibrium is to contribute nothing and free-ride, while the Pareto (socially optimal) equilibrium is for all players to contribute their entire endowments. This is what creates the social dilemma.

1.1.5 P-experiment

P-experiments have been solely used for public goods games and utilise the strategy method (Selten, 1967; see also Fischbacher, Gächter & Fehr, 2001 who first coined the term P-experiment—P for preference). These experiments elicit participants' contribution responses to different hypothetical average contributions (the social norm is the average of others' contributions) made by the other group members. For example, if initial endowments were 20 points, participants are asked how much they would contribute if the average of others was 0, 1, 2, 3, and so on up to 20 points. The experiments do not have a repeated nature and correspond closely to one shot public goods games. If participants were asked how much to contribute under ten hypothetical average contributions of others' then this is similar to taking part in ten one shot public goods games but simultaneously knowing what others are contributing. They offer the ideal means to understand what influences cooperation and what people condition it upon.

1.1.6 C-experiment

This is the direct response version of a public goods experiment. Participants take part in the public goods game C-experiment for multiple rounds (although one shot direct response games are also common; C for choice). On each round each member receives an endowment in points (the points correspond to money or prizes at an exchange rate set by the experimenter) and must decide how many of these points to allocate between their private and public account. After each period participants are told their earnings and total contributions to the public account and sometimes are also given feedback regarding the average contributions. Participants can change groups after each round (stranger design) or remain in the same group across all rounds (partner design). The decision situation payoff structure is the same as in the public goods P-experiment.

1.2 Literature Review on the Provision of Information of Others' Behaviour

Many people cooperate in public goods games under conditions where conventional economic theory predicts that rational utility-maximising individuals will, if they lack other-regarding preferences, contribute nothing.

In a typical four-player public goods P-experiment, around 50% of players can be characterised as conditional cooperators — they increase their contributions as the mean of others' contributions is increased (e.g., Fischbacher, Gächter & Fehr, 2001). More recently, reanalysing the data from six previous public goods studies and using hierarchical clustering analysis, Fallucchi, Luccasen and Turocy (2017) identified four behavioural types: zero contributors, others, conditional cooperators and weak conditional cooperators (those who match other participants' average contribution at a rate less than one). One interpretation of such behaviour is that conditional cooperators are sensitive to the social norm represented by others' average contributions and seek to conform to their perception of this social norm.

In a typical four-player public goods C-experiment, contributions are found to decline significantly between first and last rounds, exhibiting a downward trend, either in a partner or a stranger design. This effect has also been attributed to conditional cooperation, with researchers suggesting that participants try to match the mean of contributions of others in the previous round but at a lower rate (Fischbacher & Gächter, 2010). Others have also hypothesised that participants learn their dominant strategy through the rounds. This is not consistent with all findings, however: if participants re-enter a second game their contributions start higher than those in the last round of the previous game (Andreoni, 1988).

Even though the economics literature on public goods games is immense (see Chaudhuri, 2011 for a review), there are only a few studies that have tested the influence of full information about others' contributions or feedback on contribution decisions. Some of these studies have compared treatments where subjects are either given full information on individual contributions after each round of the game or information on aggregate contributions (total contributions and/or average is presented) made by the group. Weimann (1994) and Bigoni and Suetens (2012) found no significant differences between these treatments: Sell and Wilson (1991) found that individual level feedback increased contributions, while Wilson and Sell (1997) found that it reduced contributions significantly. In a recent study Gächter, Kölle and Quercia (2017) found that in the aggregate feedback treatment contributions were not higher or lower compared to the individual treatment feedback. Nevertheless, when the decision frame was changed to a payoff-equivalent take game the individual treatment led to more extreme behaviour (increased free-riding and full cooperation). In the give game there was more free-riding within the individual feedback treatment compared to the aggregate in the partner design but not in the stranger design. Full cooperation within the give framing was invariant to feedback treatment for both the partner and stranger matching.

Some studies have explored the provision of information other than the average or total contributions. For example, in a study of public goods C-experiment with 4 players (Croson & Shang, 2008), it was found that partial information about a single contribution in the previous round (max, min or random contribution) affected contributions but only if the feedback mechanism (i.e. which of the three was given to participants) was not revealed. To examine evidence from the field Shang and Croson (2009) studied the effects of social comparison on public radio donations, finding a positive effect of social information on individual contributions but more specifically finding that the most influential effect on increasing donations came from social information that was drawn from the 90th to the 95th percentile of previous contributions.

Experimental studies that have explicitly examined the content of the full list of others' contributions (individual feedback) rather than merely its presence or absence compared to an aggregate feedback treatment are even fewer in number. These studies are reported below and have tested for the influence of the amount of others' contributions on aggregate and individual contributions. Their findings are consistent with effects of quantities other than the mean. Cheung (2014) studied feedback heterogeneity in a three-person public goods P-experiment, where information about others' contributions consequently contained only two entries, with ten hypothetical scenarios, where entries took values 0, 2, 4 or 6. While increasing the average feedback increased the average contributions, increasing the difference between feedback entries (which can be thought of as a proxy for standard deviation) decreased contributions. A similar effect was found in a four-person public goods P-experiment by Van den Berg, Molleman, Junikka, Puurtinen and Weissing (2015), who used ten hypothetical scenarios with entries only taking the values of 0, 10 or 20. Van den Berg et al. (2015) also found a phenomenon whereby increasing the variation of the feedback scenarios increased the variation of the contributions. Hartig, Irlenbusch and Kölle (2015)[1] studied four-person public goods games also by using a P-experiment, with 35 hypothetical scenarios and entries taking values between 0 and 20. Their results showed that increasing the standard deviation of their three-component feedback scenarios decreased contributions, but with a large variation amongst individuals, with the majority of conditional cooperators guided by others' median or average contribution and others reacting strongly to the minimum and maximum contribution of others. Croson (2007) divided subjects into groups of four and each group took part in a ten-round public goods C-experiment. A positive relationship between individuals' contributions and the aggregate contribution of others was found. When information not only about the aggregate contribution but also about the individual contributions of the other three members was revealed, the median contribution was found to be a better predictor of an individual's contribution than was the minimum or the maximum, with the proposed explanation being that participants try to match the contributions of other members.

In other experimental games besides public goods games Bilderbeck et al. (2014) used individual level feedback and found evidence for conformity to rank-based social norms in a three-person repeated harvesting game (although inconsistent use of social norms was found for serotonin-depleted participants). In Messick et al. (1983) participants were placed in 3x2 conditions during a six-person repeated harvesting game. The first variable corresponded to overuse, underuse and optimal use feedback (this was fake feedback and roughly corresponded to the mean of extractions); the second two corresponded to low and high variance. Their results showed that high variance only increased extractions compared to the low variance treatment in the underuse condition, otherwise there was no difference. The overuse and optimal use conditions irrespective of low or high variance showed a constant pattern in decreasing extractions across rounds. Measuring giving behaviour in a three-person dictator game

[1] Cheung (2014), Van den Berg et al. (2015) and Hartig et al. (2015) developed their designs independently of our own.

Panchanathan, Frankenhuis, and Silk (2013) found on average no significant peer effects (the extent to which a behaviour of a peer affects one's giving). Nevertheless, there was a considerable amount of heterogeneity with some dictators being positively influenced by their peer, others decreasing their giving if the peer gave more and some remaining unaffected. Therefore, it was proposed that participants may have perceived differently what constitutes a norm.

1.3 Traditional and Rank-Based Approach of Context Effects in Psychology and Economics

The aggregate mechanism with which many early models in psychology incorporated the effect of context was through comparison with a fixed reference point. For example, in the psychophysics, price perception and wage satisfaction literature context effects on judgements have been modelled through a single neutral reference point, typically the mean or other average of a contextual distribution (Helson, 1947, Helson, 1964a; Helson, 1964b; Briesch, Krishnamurthi, Mazumdar, & Raj, 1997; Mazumdar, Raj, & Sinha, 2005, Clark & Oswald, 1996). Indeed, most of the traditional public goods games studies use the mean of other's contributions as feedback and information regarding the social norm and compare contributions relative to a mean. This perspective, in which the response to a single reference point (the mean) is the primary observational criterion, can be compared to the aforementioned theories in which judgements are made relative to a single neutral reference point.

An influential step in moving beyond single reference point comparison is Parducci's Range Frequency theory (RFT) (1965). According to RFT many comparison stimuli affect the judgement of a particular item through a weighted sum of range and frequency (rank) components, which measure an item's range and relative rank position in the set of contextual stimuli. More recently the Decision by Sampling (DbS) (Stewart, Chater, & Brown, 2006) framework has hypothesised that the range component of RFT can be accounted for by rank effects. In DbS an individual initially retrieves a contextual sample from memory. A judgement is then made by computing the relative rank position of some quantity within a comparison set.

This work, showing that the subjective magnitudes or judgments of simple psychophysical quantities such as size and weight (e.g., Parducci, Calfee, Marshall, & Davidson, 1960; Parducci & Perrett, 1971; Riskey, Parducci, & Beauchamp, 1979), quantities as diverse as fairness (Mellers, 1982, 1986) and prices (Niedrich, Sharma, & Wedell, 2001; Niedrich, Weathers, Hill, & Bell, 2009) are determined partly by the relative ranked position they occupy within a comparison context has led to the idea that people may be sensitive to the ranked position that they occupy within an observed or inferred social distribution. Indeed, this idea has seen much currency within psychology, but the idea has typically not been translated into models within economics.[2] Effects of perceived relative rank are found within both economic and psychological domains. For example, people's anticipated and experienced satisfaction with a wage are both related to how the wage ranks within a social context (Brown, Gardner, Oswald, & Qian, 2008) and attitudes to anticipated debt are similarly determined by the ranked position of the anticipated debt relative to that of others (Aldrovandi, Wood, Maltby, & Brown, 2015). Rank of income, rather than income per se, determines satisfaction with that income (Boyce, Brown, & Moore, 2010; Smith, Diener, & Wedell, 1989). Melrose, Brown and Wood (2012) showed that an individual's self-evaluation of symptom severity in depression and anxiety was driven by rank-based comparisons with symptoms of others. Wood, Boyce, Moore and Brown (2012) found that the relative rank of an individual's income can predict risks of current and future mental distress. Rank-based decision making has also been found in social judgments (Wood, Brown, & Maltby, 2012) relating to the perception of alcohol consumption.

1.4 Applying a Rank-Based Approach to Public Goods Games

As discussed most of public goods game studies use the mean of other's contributions as feedback and information regarding the social norm. It is clear though that other attributes of others' behaviour may also affect contributions. As a simple example consider the following situation. Suppose there are two pairs of contribution profiles regarding the contributions of others in a public goods game: 9, 10, 11 and 1, 10, 19. In both cases the mean of the feedback is ten. However, there are many other

[2] Although see Hopkins and Kornienko (2004).

aspects of these profiles that one can take into account that would give different contributions. The natural way to study these effects is with P-experiments and indeed this what three studies have done so far (all of which came out while we were planning or conducting ours: Cheung, 2014; Van den Berg et al., 2015; Hartig et al., 2015). These studies have found that feedback heterogeneity (differences between entries as proxy for standard deviation) may also have an effect additional to that of the mean. These studies were however not able to offer a model of individual behaviour or what may underpin their findings.

While the existence of context and rank effects on judgement, preferences for risk, and life satisfaction (e.g., Parducci, 1965; Brown et al., 2008, 2015), along with the psychological processes that underpin them, is well documented, there is a paucity of evidence regarding the influence of contextual stimuli on cooperative behaviour and the relevant psychological mechanisms. On the basis of the typical public goods games designs it is not possible to determine whether participants are concerned with how their own contribution relates to the mean of others' contributions or with how their contribution ranks within an inferred distribution of others' contributions (the "social norm"). However, intuition and experimental evidence in psychology discussed previously suggests that individuals may show rank-sensitivity in their contributions, judging themselves in terms of their ranked position within a social sample (rather than comparing to a mean). For example, consider again the two sets of others' contributions in a public goods game: 9, 10, 11 and 1, 10, 19. In both cases the mean of the feedback is ten. It seems intuitive, however, that a fourth contribution of 8 will seem to conform less well to the social norm in the former case (where it is 2 standard deviations away from the mean of the best fitting normal distribution to others' contributions) than in the second case (where the contribution of 8 is 0.22 standard deviations away from the mean).

We illustrate the key idea of our model in Figure 1.1 below, which shows two hypothetical sets of other players' contributions in a public goods game. Each filled circle represents a contribution (we assume 10 players, and hence 9 "others"). The superimposed normal distributions represent the different "social norms" that the observer (the tenth player) might infer in each case. Thus, if the observer made a contribution of 7 it would lie at the 14th percentile in distribution A, but at the 29th percentile in distribution B (vertical dashed lines). Conversely, we might hypothesise that the observer has a preference for making a contribution that lies at the 80th percentile of the social norm. This would equate to a contribution of 12.3 in set A, but to a contribution of 14.6 in set B (vertical solid lines). The model that we develop below is essentially a generalisation of this approach.

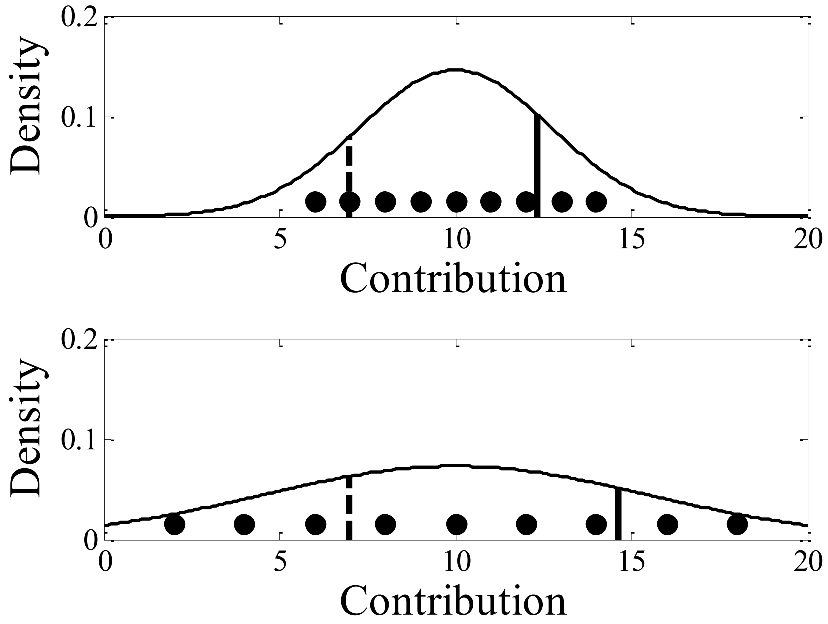


Figure 1.1. Two hypothetical sets of other players' contributions to a public goods game. Each filled circle represents a contribution. The superimposed normal distributions represent the different "social norms" that the observer (the tenth player) might infer. The dashed lines show that a contribution of 7 lies at the 14th percentile in distribution A, but at the 29th percentile in distribution B. Conversely, the 80th percentile equates to a contribution of 12.3 in set A, but to a contribution of 14.6 in set B (vertical solid lines).

1.5 Present Work

Here we adapt and extend the rank-based approach to social norms in public goods games. Informed by our pilot studies (see section 1.6), we develop and test a model of rank effects within social norms in a public goods P-experiment setting with groups of ten players.[3] In our model players are sensitive to the relative ranked position of

[3] The large majority of public goods experiments use groups of at most five players. Only a few studies have used groups of ten players, e.g., Isaac & Walker (1988); Isaac, Walker, & Williams (1994); Carpenter (2007). The effect of group size in social dilemmas has been recently investigated by Diederich, Goeschl and Waichman (2016) who found that increasing group size by a factor of 10 increased the size of the public good by 10, with cooperators cooperating more in larger groups and cooperation ratings declining more slowly in larger than smaller groups. Free-riding remained unchanged with group size.

their own contributions within a social norm. The social norm is assumed to be a continuous distribution that is inferred from the contributions of other players. Following Fischbacher, Gächter and Fehr (2001), our study used a variant of the strategy method (Selten, 1967) to create a modified public goods game P-experiment. Individuals are presented with a series of contribution profiles consisting of others' contributions and asked how much they would contribute in each case. Our use of ten-player groups allows for large variation in the profiles of others' contributions and thus provides the opportunity to explore the effects of subtle characteristics of these profiles. This is in contrast to contributions given by a smaller group, containing for example four individuals as in Van den Berg et al. (2015) and Hartig et al. (2015), where the feedback would possess only 3 entries and so have only three variables—specifying the mean in such a profile leaves only two other parameters to vary.[4] Also, in these studies the contribution profiles are shown simultaneously to participants on a single screen with ascending order of the mean. We use different screens for each of our constructed contribution profiles, the order of which was randomised for each participant.

In the present work we do not make an attempt to explain or understand social preferences or why people want to be fair or care for social welfare, but to describe and understand specific strategies in the game. What we attempt is to show that the social norm can be more than just the mean of others' contributions and that participants may have a preference for occupying a ranked position in the profiles of others' contributions. We do not posit that rank-based cooperators care about fairness or inequality aversion (although a rank model could potentially nest such preferences).

Our first key finding is that the first three moments of the distribution of others' contributions (the mean, variance and skewness) all have significant effects on contributions. Increasing the standard deviation of the contribution profiles increases the standard deviation of responses. The results indicate that comparison with a single

[4] The contribution profiles contain information on contribution decisions and not individual earnings. We do not make any attempts to generalise our hypotheses and results from contribution space to earning space. Analyzing the data from 71 studies that used the voluntary contributions mechanism, Fiala and Suetens (2017) showed that transparency about individual earnings reduced contributions while transparency about individual choices increased contributions but to a lesser extent.

reference point cannot accurately model the influence of feedback on individual's decision making.

We then develop a model according to which an individual uses rank-based decision making to place themselves within an inferred 'social norm' (a continuous distribution inferred from the contribution profile). We assume that the selected contribution reflects a utility-maximising compromise between the utility gained from free-riding and the individual's desire to conform to a perceived socially appropriate value ν_i. We fit the rank-based model to individual data and compare its fit to that of two other models – a "constant contribution" model and a "mean relative" model according to which participants are assumed to have a concern with how their contribution relates to the mean of others' contributions. Most participants' (85%) data were significantly better fit for one model than by another (i.e., there were "clear winners"). Of these clear winners only 24% were unaffected by the context of others' contributions (6.2% were altruists; 10.9% were free riders). The majority of participants (45%) were best described by the rank model, while 31% of participants were best described by the mean relative model.

We also conducted and report on a public goods C-experiment study which is not as controlled as the P-experiment, results of which are reported after the discussion of the P-experiment.

1.6 Pilot Studies that Informed Present Studies

Many studies treat the public goods and the common pool resource dilemma games played over multiple rounds or one-shot as equivalent (Ledyard, 1994), only acknowledging the framing and wording as the main difference between them, with resource games being reduced to take-some games and public goods games to give-some games. This is not true as the harvesting common pool resource game, for example, has a dynamic layer. Actions and payoffs in one round influence actions and payoffs in the next. In such a game, participants have access to a common resource pool filled with money, points or fish which they have to harvest in a series of rounds. After each round the resource gets replenished by a factor. If the resource collapses from over-depletion there will be no points to go around. Therefore, it is individually rational to extract as much as possible but if everyone is behaving in such a way the resource will collapse because the replenishment rate will be smaller than the total extractions and there would be no opportunities for future gains. A typical structure involves the following: Group of $N = 4$ players. In each round, players must choose to harvest from 0-20 points from a common shared resource. The size of the resource, R, is replenished each round by a fixed factor $f = 5\%$ as

$$R_{t+1} = (1+f)(R_t - \sum_{i=1}^{N} x_i), \qquad (2)$$

where x_i, is the amount player *i* harvests from the common resource. What is extracted by players cannot be returned back to the pool in future rounds.

Motivated by Bilderbeck et al. (2014), who found preliminary evidence for conformity to rank-based social norms in a repeated harvesting game (although inconsistent use of social norms was found for serotonin-depleted participants) using three-person groups, we focused our efforts on expanding their work and generalising their findings to bigger groups (groups of ten) as well as groups without confederates (two out of their three group members were confederates). We first ran pilot experiments using both a direct response harvesting game C-experiment (with 46 participants) and a strategy method harvesting game P-experiment (with 200 participants) in order to test our stimuli (the extraction profiles we had created) and our software. We utilized the extractions profiles S24-S27, see 1.18.2 Appendix II. Initial results from the pilot P-experiment showed that increasing the standard deviation of others' extractions increased average extractions, even when the mean of other's extraction was constant (see Figure 1.2).

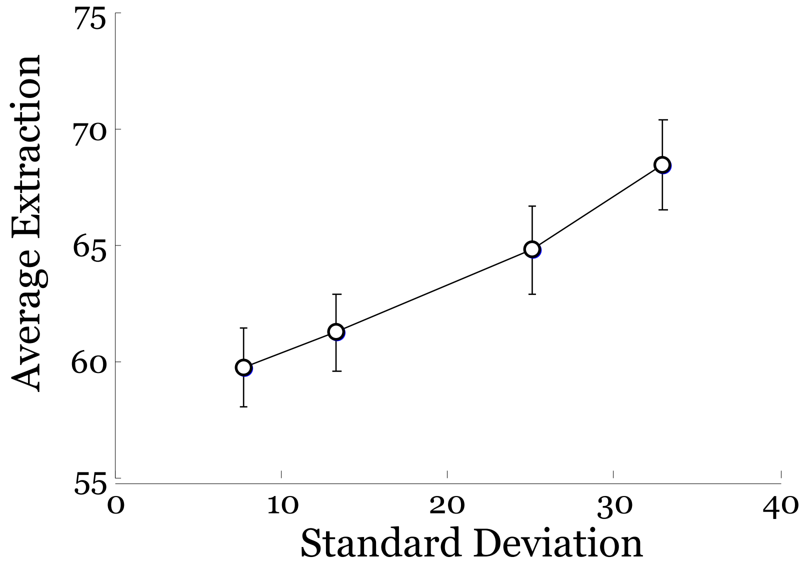


Figure 1.2. Increasing standard deviation in the extraction profiles increased average extractions, showing that not just the mean of others' extractions was influential.

Nevertheless, we only used ten extraction profiles as stimuli without controlling for symmetry within them (see 1.18.2 Appendix II which explains the idea of symmetry in the constructed profiles) and no individual-level models were developed at that point. Moreover, results from the harvesting C-experiment were not able to help us test our theories since participants depleted the resource after 4-5 rounds. More importantly we realised that it was extremely difficult for participants to understand their dominant strategy in the harvesting game and neither would we as experimenters would be able to provide them with such information through quiz questions (since the equilibrium would change depending on participants' extractions in each round and the replenishment rate). The dominant strategy and equilibrium of extracting the maximum of points allowed is easy to understand in one shot games with no multiplication factor but the harvesting game has an interior stationary Markov equilibrium where players extract too much relative to the efficient benchmark (Mailath & Samuelson, 2006) and what qualifies as "moderate" or "too much" may not have been understood by participants.

Thus, it would be difficult to communicate effectively to participants of this efficient benchmark and so their dominant and non-dominant strategies through comprehensive questions regarding the instructions of our experiment. For future work, the harvesting game can be modified to not include a dynamic layer by changing the framing of the public goods game. Instead of giving to a public good, participants can be asked to take from the public good. The structure and equilibrium would be the same in both of these give and take some games.

We now turn our focus to the current study of public goods games, which are not defined by a replenishment rate or the fact that current payoffs are not affected by previous ones. The design of the public goods P-experiment is discussed below.

1.7 P-experiment Study

1.7.1 Methods

1.7.1.1 Design

Our version of the public goods game took the following form. N individuals, labelled by an index $i \in \{1,2,\ldots,N\}$, are asked to contribute a proportion of an initial P points, x_i, to a group project. The payoff for the individual is then given by (2) where $\tau < 1$ is the marginal per capita rate of contribution. The restriction $\tau < 1$ is required for the Nash equilibrium to be zero cooperation, i.e. to create a social dilemma.

$$\pi_i = P - x_i + \tau \sum_{j=1}^{N} x_j \qquad (3)$$

We used $N = 10$, $P = 100$ and $\tau = 0.5$. Furthermore, we required x_i to be an integer between zero and a hundred, with the large range of possible contributions allowing for the creation of a wider variety of others' contributions.

1.7.1.2 Procedure

After an introduction to the study, participants read an outline of the decision situation in which the number of group members and the calculation of group project payouts were described (1.18.1 Appendix I). Participants were required to answer comprehension questions in order to confirm that they understood the rules of the task; participants who did not successfully complete these questions were not allowed to continue the experiment. The decision task itself was composed of two parts. The first part elicited each participant's choice of unconditional contribution. The second part of the task exposed the participants to 27 contribution profiles each consisting of nine numbers between zero and one hundred, presented to the participants in a random order. An example of how participants saw these 27 profiles is given below (here profile S1 is used):

"In the following you will be presented with possible sets of allocations to the group project from the 9 other group members. You will be asked to specify your preferred level of allocation to the project given the allocations of the other members. Each set of allocations is not connected to the others and you should think of them as distinct scenarios."

The allocations of the other group members are the following:

Player A	96
Player B	89
Player C	76
Player D	74
Player E	70
Player F	66
Player G	64
Player H	51
Player I	44

"How many of the 100 points do you allocate to the group project?"

We did not use the wording "invest in a group project" throughout our instructions as it could reflect a risky property. We also did not ask them for their "conditional" contribution to the project as is typically asked in P-experiments but instead asked what would they do in a situation where others contributed S [vector of others' contributions].

To understand how different contribution profiles influence contribution patterns in public goods games we designed the profiles in such a way that adoption of different simple decision strategies would lead to different patterns of contributions across the different profiles. For example, we show below that the standard deviation influences rank-based decision-making, whereas it does not influence mean-based decision-making, so the standard deviation of the profiles was varied, from 5 to 40. We varied other factors, such as the mean, range and skewness with similar motivations. The principle statistics we use to describe our contribution profiles are the mean, standard deviation and skewness and we specify a profile with the notation $\Sigma(\mu, \sigma, s)$, where μ, σ and s are the mean, standard deviation and skewness of the profile. We describe the profiles with these three statistics because they are formally independent of each other. These constructed sets of profiles might be observed rarely in genuine public goods games. To ensure that the profiles we constructed fairly represented all possibilities, we required that the average of the means was approximately 50, $\bar{\mu} \approx 50$ (actual value 50.06) and that the average of the skewness was approximately zero, $\bar{s} \approx 0$ (-0.007). Finally, we note also that the average kurtosis of the contribution profiles was also approximately zero (-0.0916). To ensure that there was no confounding effect between these variables, we constructed them such that all correlations between them were zero (see Appendix II, Table T 1.1).

No profile appears twice and within a profile we do not repeat the same contributions more than once. This ensured that a rank-based model could differentiate between all rank positions in profiles, as well as avoiding any anchoring effects coming from repeated values. As an example of our construction methods, S1 and S2 were created by holding the range and median fixed, and varying the other elements in a symmetric fashion, so that in S2 most contributions were further from the median. In this case, the profiles have the same mean, median and range, and so any decision-strategy based on these profile aspects would not yield a distinction between contributions, whereas a decision strategy, such as a rank model, would yield a distinction. Profiles S24-S27 were created by beginning with a symmetric low variance (S24), with median 50. S25-S27 were then obtained by increasing the distance of each element from the median (which was held fixed) by a constant amount. Skewed profiles were created in pairs, related by a reflection, to ensure the average skewness was zero. The result of this process was a set of 27 profiles, illustrated in Figure 1.3. This set allowed for an exploration of a variety of profile shapes without being overly burdensome for the participants. For a full discussion of these profiles please see 1.18.2 Appendix II.

1.7.1.3 Participants

The experiments were run on Amazon Mechanical Turk (MTurk). 300 individuals participated and were paid a flat fee of $3.00 along with a bonus payment which was part of the incentivisation mechanism detailed below. The payment of $3.00 was similar to the minimum US hourly wage given the average time to complete the study.[5] We incentivized decisions by randomly choosing one of each participant's responses, calculating their payoff according to the rules of the public goods game, and using the corresponding contribution profile as the other participants' contributions. This payoff was then converted into a monetary bonus payment in MTurk. If the unconditional response was chosen, then unconditional responses of 9 other randomly chosen participants were used to compute the payoff. After the experiment was completed, one of the 27 profiles and the associated unconditional response was randomly chosen for each individual and used to determine their bonus payment.

After the completion of the P-experiment participants were asked to complete some psychological scales relating to personality and conservatism. Detailed results of the analysis between contributions and these measurements can be found in 1.18.3 Appendix III.

[5] It has been consistently found that MTurk results are reliable and of high quality in behavioral experiments (Buhrmester, Kwang, & Gosling, 2011) including studies in the areas of both judgment and decision making (Paolacci, Chandler, & Ipeirotis, 2010) and economic games (Amir, Rand, & Gal, 2012; Arechar, Gächter, & Molleman, 2018).

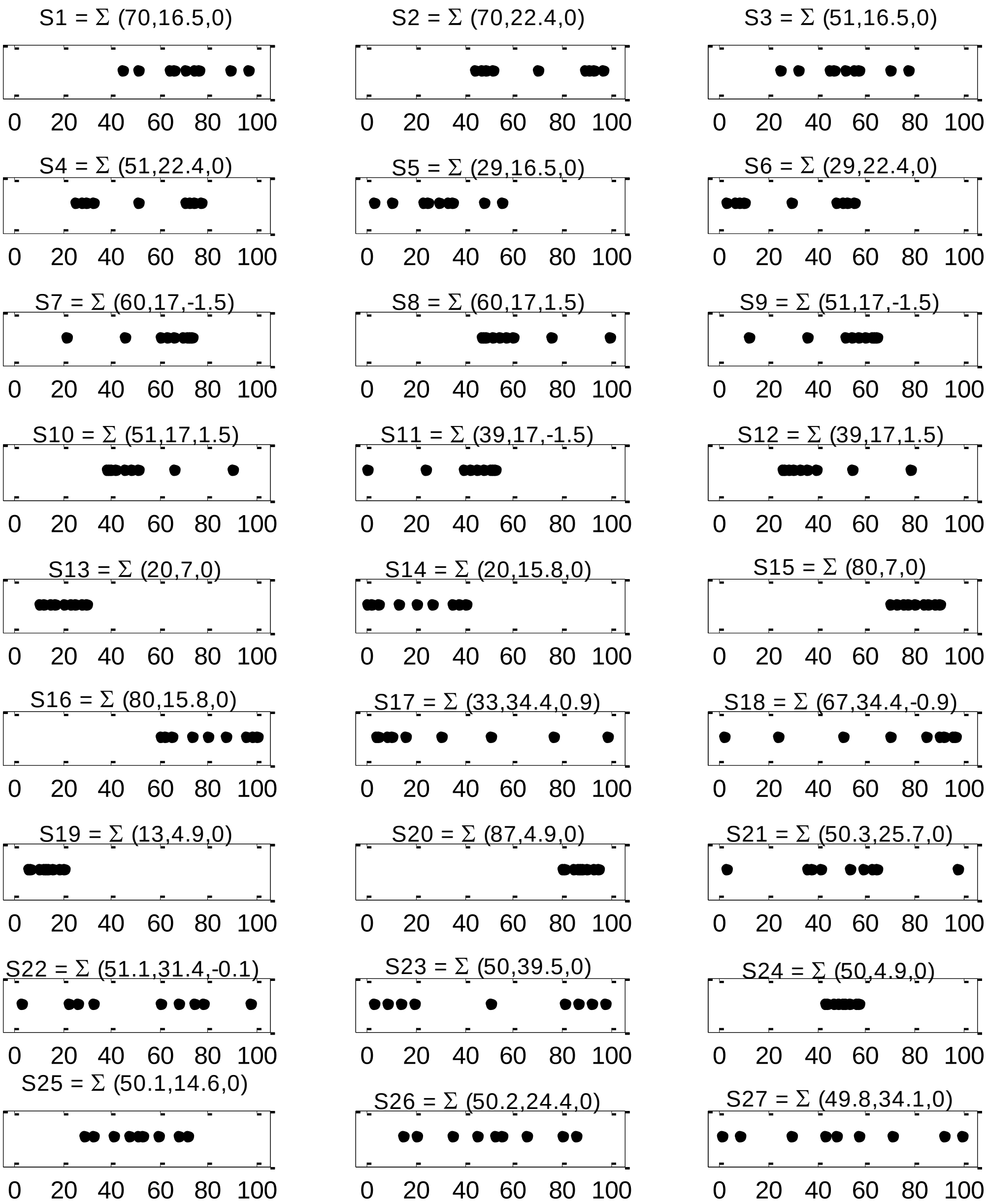


Figure 1.3. Diagrams of contribution profiles used. For a full discussion of these profiles please see section 1.18.2 Appendix II.

1.8 Aggregate Results

We first looked for evidence of conditional cooperation in the aggregate data. Of the 300 participants, 47 gave exactly the same contribution for each of the 27 contribution profiles they saw. These included 16 altruists (always contributed 100), 28 free-riders (always contributed zero) and 3 participants who always contributed other constant amounts. As these participants were not sensitive to others' contributions, and indeed we did not expect everyone to be sensitive to or willing to consider what others are contributing, we excluded them from the following analysis as the theory would not apply to them. Figure 1.4 shows the relationship between the mean of the contribution profile and the mean of the contributions made by participants, excluding those whose contributions were exactly the same for each profile (47 individuals). The correlation r between the two is 0.99, $p < .001$, indicating a very strong linear relationship. This is notable because the participants were not shown the mean of the profile during the experiment. Fitting a least squares line to the data gives a slope of 0.80 and an intercept of 11.51, showing strong conditional cooperation. It is important to note that, as we discuss in section 1.10.1, this result is consistent with either a rank-based or a mean-based model of sensitivity to social norms.

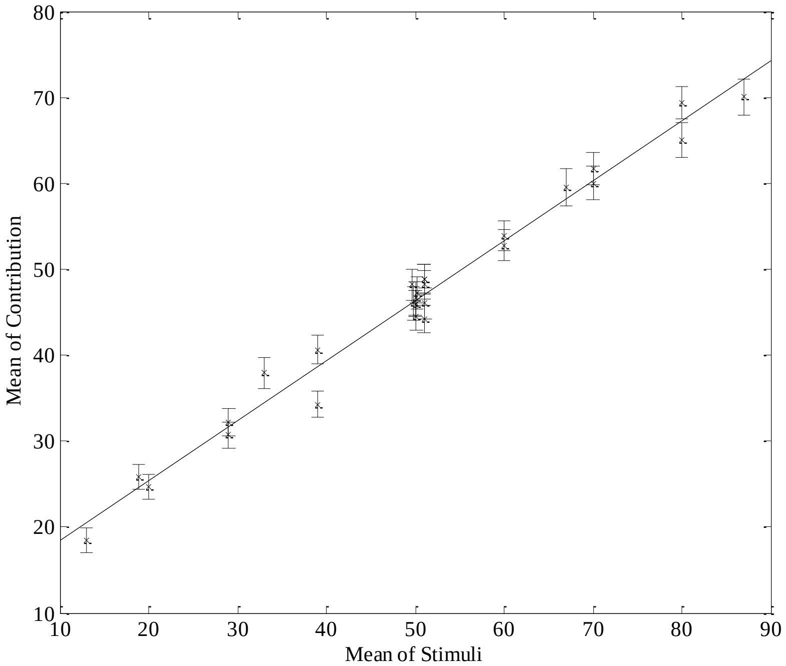


Figure 1.4. Mean of contributions from individuals whose contribution varied vs. mean of contribution profiles, the stimuli participants saw in random order.

We ran a random effects regression, with the mean, standard deviation, skewness and kurtosis of the profiles as independent variables and the individual as a random effect.

Here we also selected only those individuals whose contributions were not constant (253 out of 300, or 84%)[6]. The regression took the form

$$x_{ij} = \beta_0 + \beta_1 \mu_j + \beta_2\sigma_j + \beta_3 s_j + \beta_4\kappa_j + \varepsilon_i + \varepsilon_{ij,} \tag{4}$$

where x_{ij} is the contribution of individual i for a given contribution profile j, μ is the mean, σ is the standard deviation, s is the skewness and κ the kurtosis of the contribution profile. The coefficients β_i describe the aggregate effects of mean, standard deviation, skewness and kurtosis on individuals' contributions. We use the standard forms for skewness and kurtosis, which control whether the distribution has a heavier tail at high contributions (positive skewness) or low contributions (negative skewness) and whether the tails of the distribution are heavier than a normal distribution (positive kurtosis) or smaller (negative kurtosis). The results of this regression are given below.

*Table 1.1. Aggregate random effects regression results for contributions depending on the first four moments of the contribution profiles. Significance levels: ***<.001.*

	Estimate	Standard Error	df	t value
(Intercept)	7.86***	1.47	506	5.37
Mean	0.80***	0.01	6574	77.41
Standard Deviation	0.19***	0.02	6574	9.15
Skewness	1.11***	0.25	6574	4.45
Kurtosis	-0.04	0.18	6574	-0.24
Obs:253x27 =6831				

The first three moments, the mean, standard deviation and skewness, all have significant aggregate effects on contributions. Since the kurtosis did not display a significant aggregate effect we focus on the first three moments, and will not discuss

[6] Running the regression for all individuals or those indicated as conditional cooperators (through individual model fitting as discussed in 1.12) yielded similar results.

it further. As expected from Figure 1.4, the mean had a significant effect on contributions, with a coefficient of 0.8, so that an increase in the mean of the profile by 1 increased aggregate contributions by 0.8, demonstrating a strong effect of conditional cooperation. Despite the considerably different structure of the games (number of players, range of contributions etc.), this value is close to the values found by Cheung (2014) and Hartig et al. (2015). The standard deviation entered with a positive aggregate effect of 0.19. The effect has the opposite sign to that found by Cheung (2014) and Hartig et al. (2015) and suggests that increasing the standard deviation of the profiles increases contributions. Finally, we note that the skewness, which controls whether the distribution has a heavier tail on the right of the distribution (positive skew) or heavier tail on the left of the distribution (negative skew), had a positive aggregate coefficient of 1.1, suggesting that a heavy tail on the right of the distribution increased contributions. We note that this effect is independent of those for the mean and standard deviation and corresponds to the effect of varying the skewness while keeping these other measures constant. Aggregate effects do not tell the whole story here, and there is a large heterogeneity in the distribution of regression coefficients for individuals.

Figure 1.5 shows the distribution of regression coefficients found for individual regressions run for each participant (excluding those who contributed a constant amount). The regressions were formulated as

$$x_{ij} = \beta_{0i} + \beta_{1i}\,\mu_j + \beta_{2i}\sigma_j + \beta_{3i}s_j + \beta_{4i}\kappa_j + \varepsilon_i, \tag{5}$$

where the symbols have the same meaning as previously, but each individual has their own set of regression coefficients. The distribution of coefficients for the mean in Figure 1.5 shows a positive coefficient for almost all individuals. Conversely, the individual coefficients for standard deviation show a high degree of heterogeneity, with a large proportion being negative, as opposed to coefficients on the mean (where almost all coefficients are positive). We note, therefore, that any model of individual decision-making in these scenarios must be able to incorporate both an increase and a decrease in contributions as a function of increasing variance. As we show below, a rank-based model can account for this phenomenon (assuming that some individuals have a preference to make contributions that occupy a percentile position lower than the median of the social norm, while others' preferences are to contribute above the median). An interesting consequence, discussed in detail in 1.10.2, of this spread in individual responses to standard deviation is that increasing the standard deviation of the contribution profiles should increase the standard deviation of responses. This relationship, which can be thought of as a signature of rank-based models, is confirmed in Figure 1.6 and demonstrates a strong correlation coefficient of $r = 0.76$, $p < .001$.

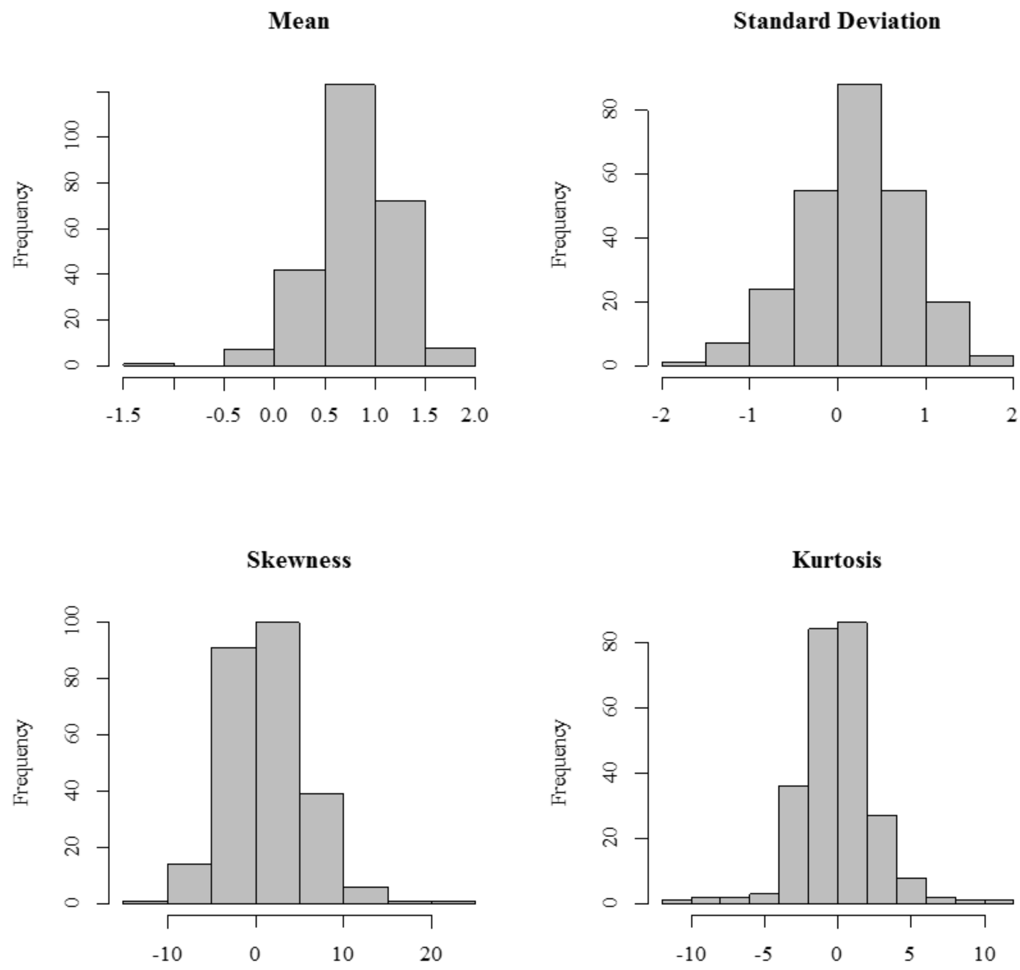


Figure 1.5. Distributions of individual regression coefficients for individuals who did not contribute a constant amount.

Finally, we note that skewness also has different effects for different individuals. We will have relatively little to say about these subtle effects of skewness, as neither the rank-based model nor the mean-based model of social norms that we develop makes predictions about the effects of skewness[7].

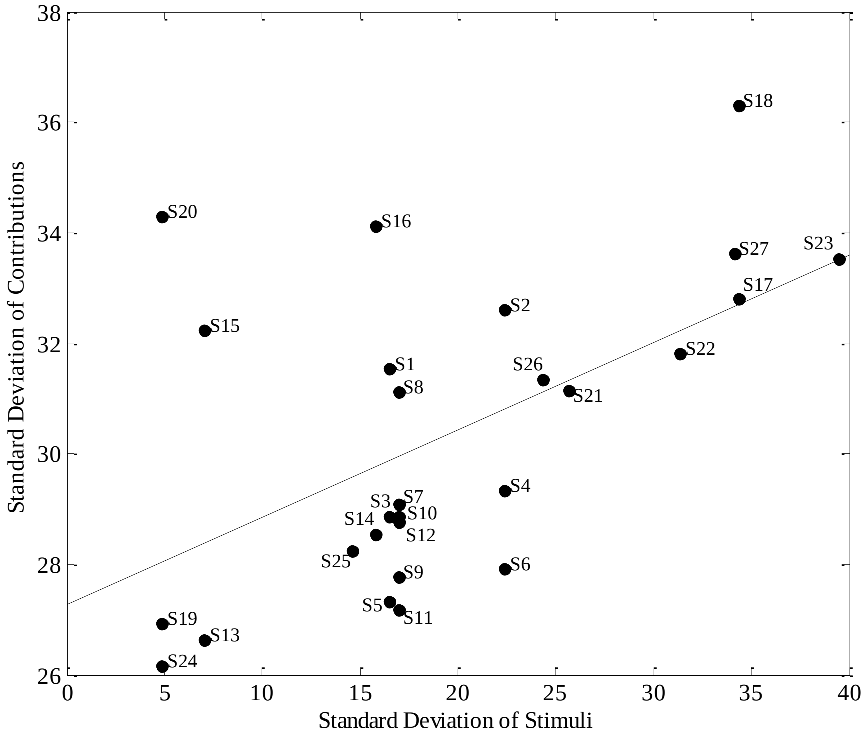


Figure 1.6. Standard deviation of contribution profiles versus standard deviation of contributions of individuals who did not contribute a constant amount.

1.9 Specification of Models

In this section we formulate our rank-based model and several alternatives with which we compare it. The purpose of this model is to capture the effects of social comparison and its interplay with players' incentives to earn as much from the game as possible.

---

[7] We note however that some accounts in the psychological literature (Brown & Matthews, 2011; Brown, Wood, Ogden, & Maltby, 2015) argue that, with additional assumptions about individuals' priors, purely rank-based processes can give rise to effects of skewness.

We assume a competition between the utility gained from free-riding and the individual's desire to conform to a perceived socially appropriate value ν_i. For the purposes of fitting the model we choose the functional form

$$U_i = P - x_i + \tau \sum_{j=1}^{N} x_j - \gamma_i (x_i - \nu_i)^2, \tag{6}$$

where P is the initial endowment (equal to 100), x_i is the contribution of the i^{th} individual, $\tau < 1$ is the marginal per capita rate of return in the public goods game, ν_i is the socially appropriate value to which the individual would like to conform and $\gamma_i \geq 0$ measures the degree to which the individual i is influenced by the social context, which we call 'norm sensitivity'[8]. The expected contribution, x_i^*, is then found by maximizing U_i, which requires $dU_i/dx_i = 0$ and $d^2U_i/dx_i^2 < 0$. This gives a solution for the expected contribution as

$$x_i^* = \nu_i - \frac{(1 - \tau)}{2\gamma_i}. \tag{7}$$

That is, as the norm sensitivity decreases, the second term in the above expression gets larger, and the individual's behavior tends more towards the Nash equilibrium value of 0 contribution (which is predicted if all individuals have $\gamma_i = 0$). As the norm sensitivity becomes very large, i's contribution tends towards their perception of ν_i. In terms of this model free riders, those who contribute zero, can be thought of as having no norm sensitivity ($\gamma_i = 0$).

The key aspect of the model is the choice of ν_i, which will in principle depend on the profiles of others' contributions. We will discuss three possibilities for this choice.

1.9.1 Constant model

The simplest possibility is a context-independent model in which ν_i is just a constant. In this case, we cannot estimate the parameters ν_i and γ_i exactly as they have a similar effect. For example, in the case of a free rider, one cannot distinguish the

[8] Also, this specification is similar to some recent models in economics that model utility from norm following (e.g. López-Pérez, 2008; Dufwenberg, Gächter, & Hennig-Schmidt, 2011; Krupka, & Weber, 2013; Gächter, Nosenzo, & Sefton, 2013).

following two cases: (a) the individual has a low norm sensitivity and so acts in her own self-interest, or (b) her perception of the social norm is such that zero is the socially preferred action and she follows this. The constant model is suited to describing free riders, altruists, and others who contribute a constant amount irrespective of the context provided by others' contributions and acts as a null model.

1.9.2 Mean Relative Model

Another simple possibility for v_i is to take the mean of the contribution profiles. In fact, we will take a constant multiplied by the mean as

$$v_i = a_i \mu. \tag{8}$$

This model represents the idea that the mean of the contribution profile is the only relevant aspect of the feedback for an individual, and can be thought of as a simple social norm model of a conditional cooperator[9]. Correspondingly, it predicts no effect of any other aspect of the profiles. The constant a_i allows for the often-observed behavior that conditional cooperators typically contribute a fraction of the mean in public goods games (Fischbacher, Gächter and Fehr, 2001).

1.9.3 Rank Model

The final choice for v_i is derived from a rank-based model of decision making. Our rank model posits that the socially appropriate value, v_i, for an individual, corresponds to a specific rank position within the contribution profile. A simple version of this model, which we term the 'discrete rank model', sets the value of v_i to be equal to the p^{th} ranked element of the contribution profile. p is a rank parameter governing the individual's behaviour and is an integer between 1 and 9 (we use the convention that a rank of 1 indicates the lowest contribution). As an example, if the contribution profile is given as (profile S1):

44	51	64	66	70	74	76	89	96

[9] We also fitted a model where individuals had a preference to be placed at a constant distance from the mean which for the majority of our participants fit worse than the mean relative model.

Then in a public goods game an individual A with rank parameter $p = 3$ would have a value for v_A of 64, and so their contribution would be equal to

$$x^*{}_A = 64 - \frac{(1-\tau)}{2\gamma_A}. \tag{9}$$

If we assume that the individual's norm sensitivity, γ_A, is large, then $x^*{}_A = 64$, or the 3rd ranked element of the feedback profile. Another individual B with rank parameter equal to 7 would have a contribution $x^*{}_B = 76$.

With this simple model we can show how a rank-based theory of contributions can account for the aggregate responses to increasing the standard deviation of the contribution profiles. To see this, take a second example where the contribution profile is given as (profile S2):

44	47	49	51	70	89	91	93	96

This profile has the same mean and skewness (as well as range and median) as the first example, but an increased standard deviation. Our two individuals, A and B, will now have contributions given by 49 and 91 respectively. In the case of individual A, increasing the standard deviation of the contribution profiles *decreased* their contribution from 64 to 49, whereas in the case of individual B, increasing the standard deviation of the profile *increased* their contribution from 76 to 91. We see that in a rank-based model, individuals have heterogeneous responses to increased standard deviation: individuals with low rank parameters will tend to decrease their contribution whereas individuals with high ranks will tend to increase their contributions. On aggregate therefore, the effect of increasing standard deviation of the profile on the mean of individuals' contributions will depend on whether more individuals have high or low rank parameters. However, the aggregate effect of increasing standard deviation of the profile on the standard deviation of individuals' contributions will always be positive.

In this simple model the rank parameter is fixed to one of nine values, and individuals' rank preferences cannot lie between these discrete values, or outside them. To allow a continuous rank preference we generalize the rank model to a continuous version. In the continuous model the discrete contribution profile is replaced by a continuous PDF, which the individual infers from the context. This PDF corresponds to the individual's perceived distribution of the actions taken by others and, as such, we refer to this PDF as the inferred social norm. Given this inferred social norm PDF, $f(x)$, the individual's socially acceptable value, v_i, then corresponds to a particular relative rank position with the social norm PDF. Mathematically, if $f(x)$ is the inferred pdf and $F(x)$ the associated CDF then v_i is given by $F^{-1}(r)$, where r is a rank parameter between 0 and 1.

At this conceptual level our model is very general, we have not specified the particular form for the PDF nor the information that determines it. Here we assume that the contribution profiles completely determine the form of the social norm PDF. The mathematical method we employ is kernel density estimation (KDE), which allows a smooth distribution to be inferred non-parametrically from a dataset (Rosenblatt, 1956; Parzen, 1962). That the method is non-parametric is key as it is important that the shape of the social norm PDF, and thus properties such as skewness and standard deviation, are inherited directly from the contribution profiles, without being constrained by a particular functional form. As an example, if we created the social norm PDF by fitting a normal distribution to the profiles, the result would always have zero skewness. The method works by placing a chosen function (the kernel) centered over each point in the contribution profiles and aggregating the results. We choose the normal PDF as our kernel function and so given a contribution profile $D = (d_1, d_2, \dots, d_n)$ of contributions by other participants, the inferred social norm PDF is given by:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} N\big(x; d_{i,}, b\big). \tag{10}$$

$N(x; d_i, b)$ is a normal PDF with mean d_i and standard deviation b. The parameter b, the bandwidth, controls the form of the PDF that results from this procedure. Since the kernel PDF is created by placing a Gaussian of width b over each point in the contribution profiles and then averaging, as the bandwidth increases the smoothness of the resulting PDF also increases. This is illustrated in Figure 1.7, where we show a series of diagrams describing this effect. When the bandwidth is small, approximately 0, the resulting PDF consists of a series of spikes centered over each point in the contribution profiles. As the bandwidth increases, the PDF becomes smoother and the modality of the distribution decreases until it becomes a uniform distribution over the allowed range of contributions.

The effect of the bandwidth on the contributions of individuals follows from these properties. To understand these effects, we will assume a high social sensitivity, so that the individual's contribution is simply given by v_i. On an intuitive level, decreasing the bandwidth increases the tendency of an individual to contribute a value that is close to one of the contributions they observe. Figure 1.8 shows an individual's contribution as a function of their rank parameter for differing values of the bandwidth. We see that for $b = 0.1$, the majority of rank parameters lead to a contribution that is equal to a value in the contribution profiles; this corresponds to the plateaus in Figure 1.8. As the bandwidth increases to $b = 10$ the graph becomes smoother, and the majority of rank parameters lead to contributions that lie between values in the contribution profiles.

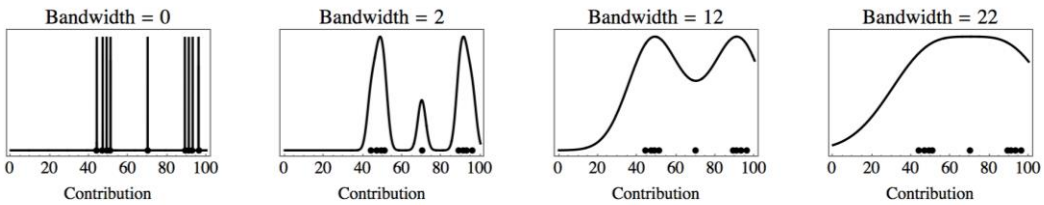


Figure 1.7. Effect of varying the bandwidth on the inferred social norm PDF. When the bandwidth is small the inferred PDF is a series of spikes over each contribution in the contribution profiles. As the bandwidth increases, the modality of the inferred PDF decreases until it becomes a unimodal distribution.

The behavior of the continuous rank model when the bandwidth is either very small (equal to 0) or very large (approaching infinity) is simple, and replicates the behavior of other models, the discrete rank model and the constant model respectively. When the bandwidth is zero we obtain the discrete rank model that was introduced at the beginning of this section, since all the probability in the KDE PDF is concentrated over the values in the contribution profiles. When the bandwidth is very large the KDE PDF is the uniform distribution between 0 and 100, and so the contribution is given by 100 (the range of possible contributions) times the rank parameter.

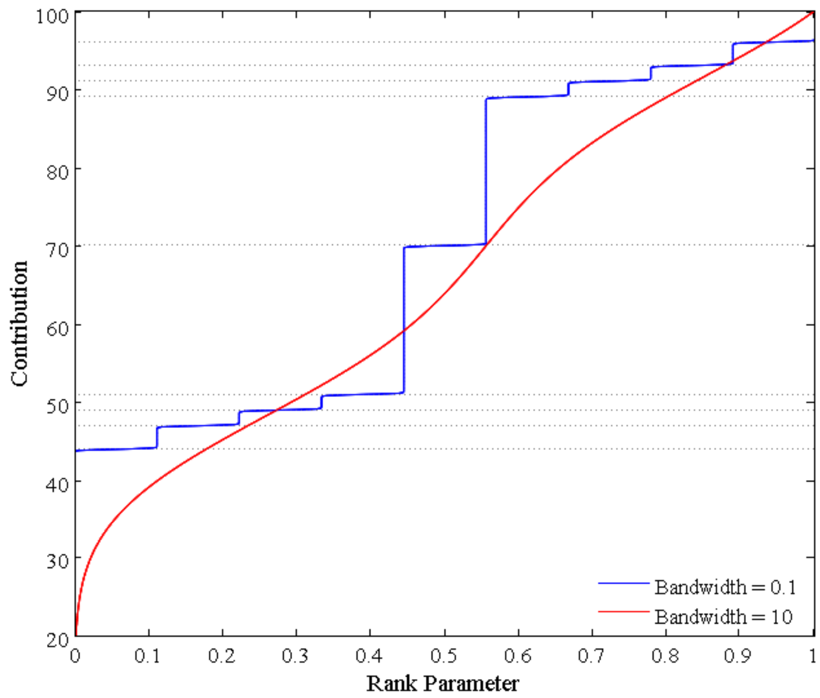


Figure 1.8. Contribution versus rank parameter for two different bandwidths, 0.1 and 10, for the contribution profile S2. When the bandwidth is small (0.1) most rank parameters correspond to contributions close to values in S2 (denoted by the horizontal dashed lines). As the bandwidth becomes larger (10) the curve becomes smoother, and contributions vary smoothly between the values from the profile.

1.10 Model Predictions

We will now show how, on both an aggregate and an individual level, the rank model predicts responses to changes in the structure of the contribution profiles. We will focus on the effects of the mean and the standard deviation.

1.10.1 Mean

Consider a profile that produces a particular set of contributions from individuals, whom we suppose make their decision according to the model specified above. We can now ask what contributions we will obtain if the mean of the profile is increased, which we do by increasing the value of all entries in the profile by a fixed amount, say

10. Then, the pdf that is inferred from the profile will just be shifted by 10, and so the contribution associated to any given rank parameter will also increase by 10 (see Figure 1.9). Thus, the model predicts linear increase in contributions as a function of the mean of the feedback on both an individual and aggregate level. This linear relation is precisely what we observe and indeed is the behavioral archetype of the conditional cooperator as defined by Fischbacher, Gächter and Fehr (2001).

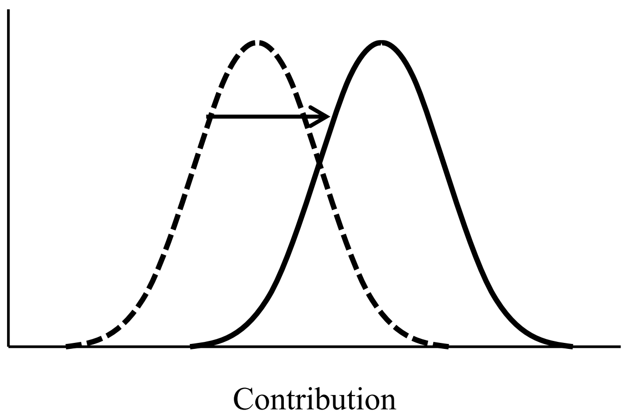


Figure 1.9. Increasing the mean of the inferred social norm pdf can be accomplished by shifting the entire distribution to the right. As a result, the contribution corresponding to each rank parameter is also shifted by this same amount.

1.10.2 Standard Deviation

We now turn to the predicted effect of increasing the standard deviation of the feedback on contributions. In contrast with the previous case of increasing the mean, the rank model predicts heterogeneous individual responses to changing the standard deviation of the contribution profile. Increasing the standard deviation of the contribution profile increases the standard deviation of the inferred social norm pdf, which can be thought of as a widening of the inferred pdf, as illustrated in Figure 1.10. Increasing the standard deviation of a distribution without affecting other moments such as the mean or skewness can be accomplished by pushing each element in the profile away from the mean. The result of this is that individuals with low rank parameters will decrease their contribution, shown as α, and individuals with high rank parameters will increase their contribution, shown as γ. The crossover point between the two effects is at the value of the rank parameter corresponding to the median of the distribution which will typically be around 0.5, shown as β. Standard deviation therefore has a markedly different effect to the mean: increasing standard deviation will either increase or decrease the contribution of an individual depending on the value of their rank parameter.

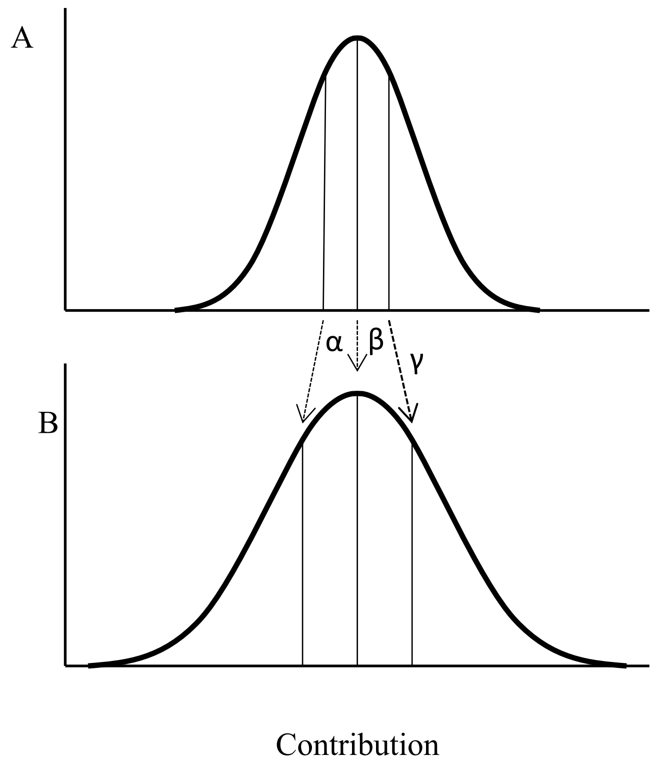


Figure 1.10. Increasing the standard deviation of the contribution profile can either decrease an individual's contribution (α), increase their contribution (β) or result in no change (γ), depending on the rank parameter. As a result, the standard deviation of individuals' contributions will increase on aggregate.

Of interest here is therefore the aggregate effect. It is clear from this that our model unequivocally predicts that increasing the standard deviation will increase the standard deviation of the contributions. This is precisely what we observe, as shown in Figure 1.6. The effect of increasing the standard deviation on the mean of contributions is more subtle. The direction of the aggregate effect will depend on the spread of rank parameters within the population. If the average rank parameter is low then we expect the aggregate effect to be negative. If the converse is true we expect the aggregate effect to be positive. Certainly, however, if there are significant numbers of individuals with both high and low rank parameters we expect the aggregate effect to be small. Our data shows a small positive aggregate effect of the standard deviation. Interestingly Cheung (2014) and Hartig et al. (2015) found an aggregate effect of standard deviation in the opposite direction. These findings are not necessarily inconsistent however, since, as discussed, the aggregate effect may vary. Depending on the sample and the individual preferences, high variance can lead to different or mixed results.

1.11 Model Implementation

To compare the rank-based model with the two other candidates (mean-based and constant), we fit the models to the data from individual participants. This is done through a maximum likelihood framework (as in Lewandowsky & Farrell, 2010). We take the log-likelihood for individual i to be

$$LL_i = \sum_{j=1}^{27} \log N_s\big(x(\Sigma_j; \theta_i) - c_{ij}, \sigma_i\big), \tag{11}$$

where c_{ij} is the measured contribution of individual i in contribution profile j. $x(\Sigma_j; \theta_i)$ is the predicted contribution from the model for j, with parameters θ_i. $N_s(x, \sigma_i)$ is a modified version of a normal PDF with mean 0 and standard deviation σ. It is made by discretizing the normal distribution over the integers between 0 and 100 (inclusive). All the probability above 100 is added to the value of N_s at 100. The psychological motivation for N_s is to account for the constrained intention of an individual to contribute below 0 or above 100. That is, if an individual would like to contribute 125, but cannot due to the game rules, they will instead contribute 100, so all the probability of contributing above 100 is added to 100. The equivalent process is applied to contributions below 0. Parameter values θ_i (which contains three parameters for the rank model, two for the mean model and one for the constant model) and σ_i are then found by maximizing the log-likelihood. Note that because we have used a discrete distribution, the log-likelihood is always ≤ 0.

1.12 Individual Results

Table 1.2 below shows the results of evaluating the three competing models — the rank-based model, the mean-relative model and the constant model. We fit each model for each individual using maximum likelihood estimation. Since the models contain different numbers of parameters we use AICc to compare them (following Lewandowsky & Farrell, 2010). We say that one model has a clear advantage over another if its AICc differs by two or more. It is clear that even using AICc to correct for the number of parameters the rank-based model fits a substantially higher number of people than either the constant model (the majority of the winners for the constant model are free riders and altruists) or the mean relative model.

Figure 1.11 shows the effect on the proportions of the clear winners for each of the models as the AICc threshold is varied. The rank model increases its advantage as the threshold increases, illustrating the robustness of our findings.[10]

Table 1.2. Results of model fitting. Average AICc is for all participants, not just clear winners, showing the rank model with the superior average AICc.

Model	Clear Winners	Percentages Of Total Clear Winners	Average AICc	Number of Parameters
Constant	61 (including 16 altruists and 28 free-riders)	24%	175.903	2
Mean Relative	80	31%	159.354	3
Rank	115	45%	150.983	4
Total Clear Winners	256			
Total Participants	300			

[10] The average contributions of rank and mean relative winners were 54.60 and 52.13 respectively. The average standard deviations of the contributions of rank and mean relative winners were 18.06 and 22.71 respectively. These were not found to be significantly different.

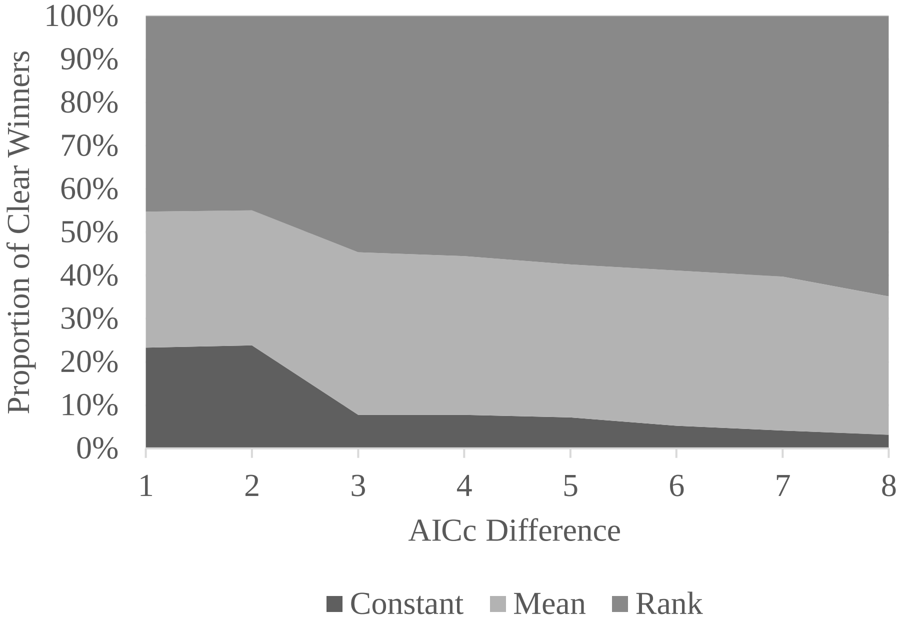


Figure 1.11. Effect on clear winner proportions as the AICc threshold is increased.

Figure 1.12 shows selected estimates of model parameters for the rank model and the mean relative model obtained through maximum likelihood estimation. The estimates of the *rank parameter*, r_i, show a peak at around 0.5. Notably, this distribution has more individuals above this modal value than below it. This is consistent with our discussion of the effect of the standard deviation, where we showed that if a majority of individuals have a rank parameter above around 0.5, then we expect a positive aggregate effect of standard deviation on contributions. The estimates of the *bandwidth*, β_i, show a monotonically decreasing shape. The mean bandwidth is 17.2, with the majority of individuals having a bandwidth below 20. We can conclude that the majority of rank winners potentially take into account some more subtle aspects of the feedback profiles. We also see the estimate for the *mean slope parameters*, α_i, in the mean relative model. We note that there is heterogeneity within the clear winners and that all the coefficients are positive. The mode of the distribution is approximately 1, indicating that these individuals tend to follow the mean of the distribution very closely.

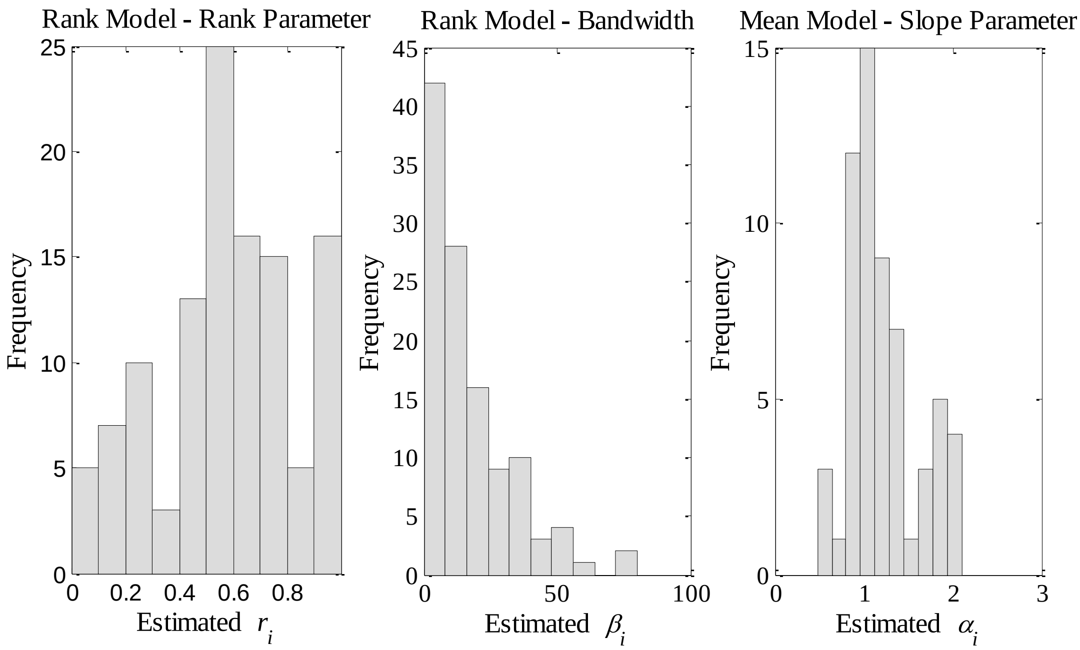


Figure 1.12. Estimated model parameters obtained through maximum likelihood. From left to right. Estimated rank parameters for rank clear winners. Bandwidth parameters for rank clear winners. Mean parameters for mean relative clear winners.

Figure 1.13 shows predicted versus observed contributions for two participants, illustrating rank-based decision making, and the advantage that the rank model has over the mean relative model for those participants. A perfect fit is given by points that lie on the line x=y. The left panel of the figure shows data from a participant with an estimated rank parameter of 0.29, with bandwidth 0.1 and $(1-\tau)/2\gamma = 0$ (i.e. a very large norm sensitivity). The corresponding parameters for the mean model give a slope of 1.01 and $(\tau - 1)/2\gamma = 12.30$.

The situation in the right panel is similar, with the participant having an estimated rank parameter of 0.05, a bandwidth of 0.1 and a high norm sensitivity, $(1-\tau)/2\gamma = 0$. In this case the parameters for the mean model are a slope of 0.99 and $(1-\tau)/2\gamma = 27.72$. In both these cases the best fit bandwidth parameter is small, 0.1, suggesting discrete rank behavior, and indeed this is what we observe, with effective discrete rank parameters of 3 and 1 respectively. It is clear from Figure 1.13 that mean-based decision making does not capture this behavior. Indeed, we could not find any examples of participants for whom the mean model produced a vastly better fit, capturing qualitatively different behavior. Indeed, from our observations it seems that there is no pattern of behavior that the mean model can display which the rank model cannot. We can explain this heuristically by observing that setting the bandwidth to a moderate value (say 20), a rank parameter of 0.5 will be close to the mean of the profile and so produce similar results to the mean model. Additional figures illustrating a better fit for the rank model can be found in Appendix IV.

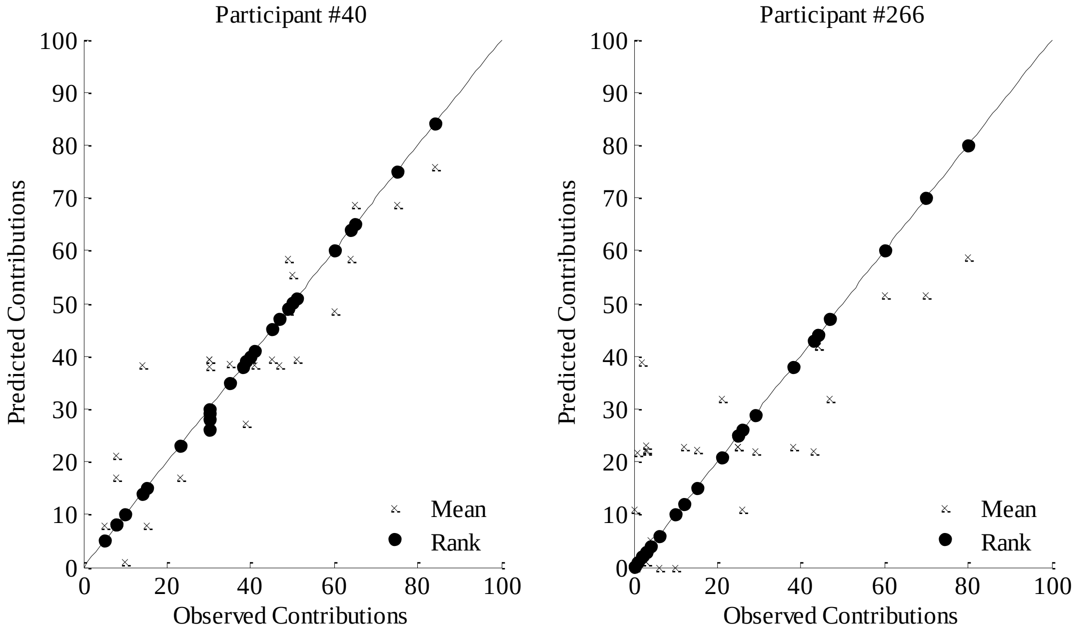


Figure 1.13. Predicted versus observed contributions for two participants, with both the mean relative and rank models, displaying a better fit for the rank-based model. In both cases the rank model perfectly fits the observed contributions, which can be seen by observing that the rank values fall on the line x=y.

1.13 Discussion P-experiment

We found strong influences of the context of others' behaviour and rank effects-- the influence of the perception of own's ranked position within a perceived social norm-- clearly distinguished.

Specifically, we found that, in a public goods P-experiment with ten players, aspects of players' contributions other than the mean had a significant effect, both at an aggregate and individual level, on contributions. This suggests that individuals care about what the average contribution is comprised of. Our study used a variant of the strategy method, allowing us to explore a full range of possible contribution profiles. We reproduced the well-established dependence of individual's contributions on the mean of others' contributions, even though this mean was not shown to participants. We found that the standard deviation increased contributions on aggregate, though the magnitude, and even the sign, of the effect varied greatly among participants. We have shown that this phenomenon is consistent with a rank-based model of decision. Further evidence for this hypothesis was given by showing that the standard deviation of the contribution profiles increases the standard deviation of the contributions observed.

On an individual level we found that a rank-based model of decision making incorporating effects of standard deviation can predict individuals' decisions better than a model in which the mean is taken to be the only aspect of the feedback to which people attend, even though it has more parameters for which it is penalized.

We note that in the economics literature there are two well-known models of social preferences, by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), which take into account social comparison, though they are typically applied to games with smaller numbers of players. However, it has been shown by both Cheung (2014) and Hartig et al. (2015) that these models do not give appropriate predictions for what we observe in our results. In Bolton and Ockenfels' model (2000) individuals prefer an equal split of payoffs between them but also prefer to earn more. As such participants will not contribute more than the average of others' contributions and what constitutes the average will not matter to them. In Fehr and Schmidt (1999) participants feel envy if they have a lower payoff than others and compassion if their payoff is higher than others. According to this model Hartig et al. (2015) showed that in bigger groups and with full information regarding individual contributions, a participant will not contribute more than the minimum of others' contributions. The behavioral predictions of these models about how people react to heterogeneity in others' individual contributions were not able to capture the effects we reported in this chapter.

Our findings are in accordance with many demonstrations within the psychological literature that individuals are sensitive to relative rank effects in for example psychophysical (Riskey et al., 1979), economic (Boyce et al., 2010; Wood et al., 2012) and social (Wood et al., 2011) contexts.

One assumption we have made in the formulation of our rank model is that comparisons between individuals are made on the basis of contributions, that is, the value of ν_i is a preferred contribution. An alternative perspective, which could form the basis of a future study, is that earnings or payoff matter. In this case, one rephrases the model such that ν_i is a preferred payoff. Indeed, it is possible that there is heterogeneity amongst individuals as to whether their preference is to think in terms of contribution space or earnings space. Experiments exploring feedback in terms of earnings in social dilemmas (Bigoni & Suetens, 2012; Nikiforakis, 2010) have found that contributions can be reduced by comparison with others' earnings rather than others' contributions and that participants exhibit a tendency to follow the best performer.

Our model, which has more general application than the experiment considered here, provides a new way to model perceived social norms. Many social norm interventions or experiments using social norm ideas (Schultz et al., 2007) provide participants with a number representing the mean level of others' activity (such as energy consumption). Our perspective suggests that such interventions might be more effective if they included information about an individual's ranked position within a social group. Two recent studies are consistent with this idea: Taylor et al. (2018) found that telling people how their drinking level ranked within the distribution of others' drinking elicited more health-related information-seeking than telling them how their behavior related to the mean level of others' drinking, and Aldrovandi, Brown, and Wood (2015) found that students' level of concern about their consumption of unhealthy foods was predicted in majority by the subjective rank of their own consumption within the social distribution. Moreover, relative willingness to pay for healthy vs unhealthy foods was similarly affected in that provision of rank-based social norm information about others' consumption of unhealthy foods ("You think 80% of students eat more chocolate than you do; in fact only 15% of students eat more chocolate than you do") was more effective than provision of mean-based social norm information ("You think the average consumption is 6 bars of chocolate; in fact it is 3).

Our model moves beyond a simple model of feedback without requiring an explicit comparison with every other individual, instead assuming the inference of a PDF. On a practical level the use of Kernel density estimation is to better estimate the rank parameters as there exists no other statistical distribution that can perfectly model all the different types of contribution profiles we gave to participants. For example, when profiles had different skew or modality it was not appropriate to use beta or Gaussian distributions. However, depending on the specific form of the profile, other types of estimation can serve the same purpose, depending on the properties of stimuli that one uses.

1.14 C-experiment Study

This experiment was designed to understand the effects of the full feedback of contributions and their content on aggregate and individual contribution decisions. Additionally, its goal was also to examine rank effects besides what is typically been studied which is the mean of the contributions in the previous round.

1.14.1 Methods

1.14.1.1 Design

Participants took part in an incentivised repeated public goods game coded using the Bonn Experimental System (BoXS). We used a partner design to achieve a higher and more stable level of contributions and to avoid the problems of large samples needed in order to bypass repeated encounters when large group of players are used (stranger design).

1.14.1.2 Participants

We recruited 56 participants using Warwick University's pool of volunteers. Each participant was promised a flat fee of £3.00 but was told that, depending on their choices, they could earn up to £10[11].

1.14.1.3 Procedure

Participants were in groups of 9 or 10 and took part in the game for 30 or 50 rounds in the Warwick Business School Behavioural Science Laboratory. Although we had overbooked the lab in order to have at least ten participants turning up in each session of the study, unfortunately a lot of them did not show up and so we had to

[11] The experiment was incentivised so that participants had a motivation to make payoff relevant decisions and attend to the decision task. We run an initial pilot version of this, testing the interactive software, with 20 Psychology students who took part in exchange for course credit. Their data are not included in this analysis.

modify our code in order to create groups of nine. In total there were six sessions: four groups of nine players who took part in the public goods game for 50 rounds and two groups of ten players who took part in the public goods game for 30 rounds. Each session lasted approximately 60 minutes.

The game had a similar payoff structure to our P-experiment. On each round each player received an endowment, P, of 100 points. Payoff in each round was given by

$$\pi_i = P - x_i + \tau \sum_{j=1}^{N} x_j, \qquad (12)$$

where x_i is the amount player *i* contributes to the public good. We set $\tau = {}^{5}/_{10}$, the marginal per capita rate of return for the public good. The payoff structure here is exactly the same as the one we used in the P-experiment making our studies comparable in terms of parameterisation of the decision situation. An example of what participants saw after a given round is given below:

"You are player "E". You contributed "70" Points. The average amount contributed is "33" Points, your payoff for this round was "196" Points."

The contributions from all players in decreasing order were:

Player "E":70

Player "D":68

Player "H":48

Player "A":42

Player "G":30

Player "C":28

Player "F":25

Player "B":10

Player "I": 5

Player "J":5

By observing themselves in the feedback, participants could infer their own relative rank position within this feedback after each round. To avoid the possible confound of hedging we did not paid participants according to their running total payoff but we randomly picked one round for each participant and paid them according to their decisions and the decisions of the other group members for that round (see 1.18.5 Appendix V for full instructions used in the C-experiment).

1.15 Results

The basic quantity typically measured in a public goods C-experiment is the average contribution over time. We show this in Figure 1.14. Figure 1.15 shows the average contributions for each of the six groups over time and showcases the large degree of heterogeneity among groups. Note the decreasing level of contributions in both figures. This is a typical pattern seen in such experiments as participants try to contribute an amount slightly below the average of the previous round and provides evidence that our design has worked well on a basic level despite the increased group size not typically used in the literature.

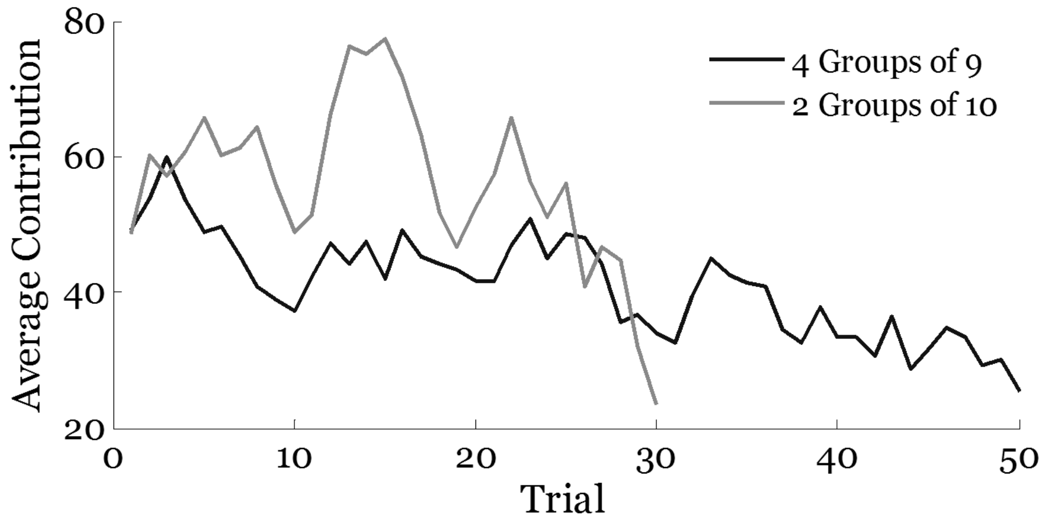


Figure 1.14. Average contributions as a function of trial number.

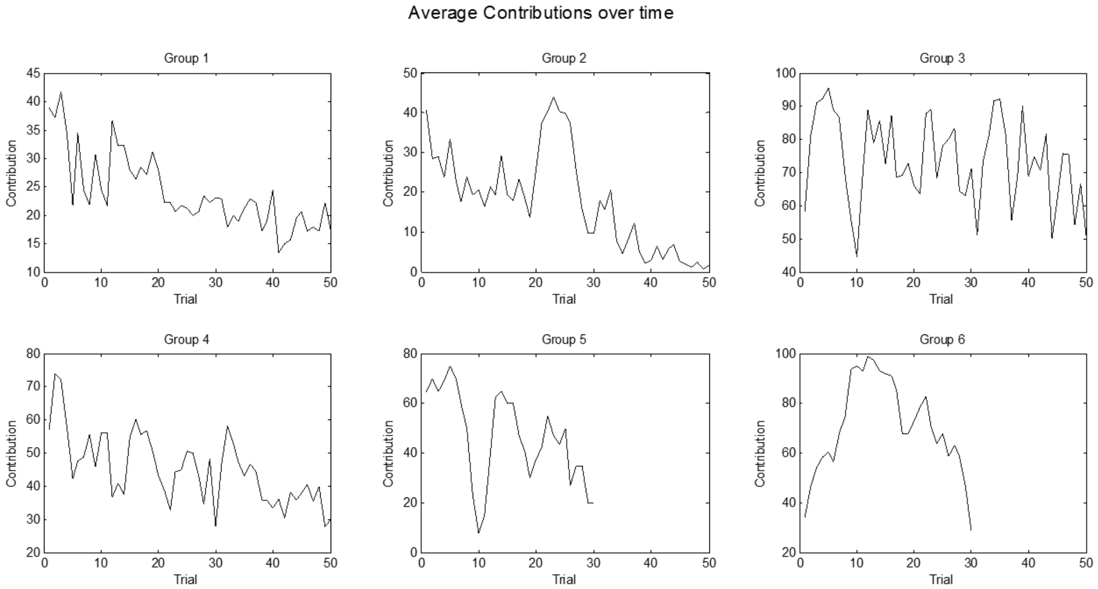


Figure 1.15. Average contributions over time for individual groups.

As in the analysis of the P-experiment we ran a random effects regression with the mean, standard deviation, skewness and kurtosis of all contributions in the previous round as independent variables, the individual as a random effect, trial number as a fixed effect and contribution in the current round as dependent variable. The regression took the form

$$C_{ij} = \beta_0 + \beta_1 \mu_{j-1} + \beta_2 \sigma_{j-1} + \beta_3 s_{j-1} + \beta_4 \kappa_{j-1} + \beta_5 j + \varepsilon_i + \varepsilon_{ij}, \qquad (13)$$

where C_{ij} is the contribution of individual i for round j, μ is the mean, σ is the standard deviation, s is the skewness and κ the kurtosis of all the contributions in the previous round. The coefficients β_i describe the aggregate effects of mean, standard deviation, skewness and kurtosis on contributions. We use the standard forms for skewness and kurtosis, which control whether the distribution has a heavier tail at high contributions (positive skewness) or low contributions (negative skewness) and whether the tails of the distribution are heavier than a normal distribution (positive kurtosis) or smaller (negative kurtosis). The results of this regression are given in Table 1.3.

Table 1.3. Aggregate random effects regression results for contributions depending on the first four moments of the previous round.

	Estimate	Standard Error	t statistic	p value
(Intercept)	22.86	5.61	4.07	<.001
Mean	0.65	0.07	8.78	<.001
Standard Deviation	-0.045	0.08	-0.60	0.55
Skewness	3.26	1.46	2.22	0.02
Kurtosis	0.13	0.24	0.56	0.58
Trial Number	-0.26	0.05	-5.39	<.001
Obs:2344				

The mean and skewness had significant aggregate effects on contributions. The mean had a positive and significant effect on contributions, with a coefficient of 0.65, so that an increase in the mean of the feedback by 1 in the previous round increases the aggregate contributions by 0.65 in the current round, demonstrating a strong effect of conditional cooperation similar to that found in Fischbacher and Gächter (2010) despite the considerably different structures of the games (number of players, range of contributions etc.). The standard deviation entered with a small negative aggregate effect but not a significant one. This is not evidence against a rank model, however — as we discussed previously in 1.10.2, if participants have heterogeneity in their rank preferences and those are symmetric around the median then one might observe no aggregate effects of standard deviation. Finally, we note that the skewness, which controls whether the feedback had a heavier tail on the right of the distribution (positive skew) or heavier tail on the left of the distribution (negative skew), has a positive coefficient of 3.3, suggesting that a heavy tail on the right in the feedback of the previous round will increase contributions in the current round. There is also a strong negative effect of trial number which confirms the effect observed in regarding the overall decreasing pattern of contributions.

The unique part of our analysis studies the effect of relative rank on participants' contribution decisions. If a participant has a preference for a relative rank position within the feedback they receive after each round then we would expect an S-shape pattern between their change in contribution from round to round and their rank position. We illustrate this idea in Figure 1.16. If at a particular round a participant's rank was lower than their preference, they increased their contribution next time, so ΔC was positive. If at a particular round a participant's rank was higher than their preference, they decreased their contribution next time, so ΔC was negative. For example, a participant has a preference to be at rank 4, they receive feedback regarding the contribution of all players in the previous round in descending order (this gave them an indirect idea regarding where their contributions ranked among others) with their contribution of 20 points corresponding to rank of 7 and not 4. This means that they should increase their contribution in the next round and so their change in contribution should be positive, $\Delta C > 0$, since they were at a lower relative rank position than they would like to (here the rank positions between 1 and 9 or 1 and 10 correspond to the highest and lowest contribution respectively). The middle in the figure represents the stable relative rank which participants try to return to, which for a given participant could be any rank position. The stable rank represents the point at the figure at which a participant would neither want to increase nor decrease their contribution.

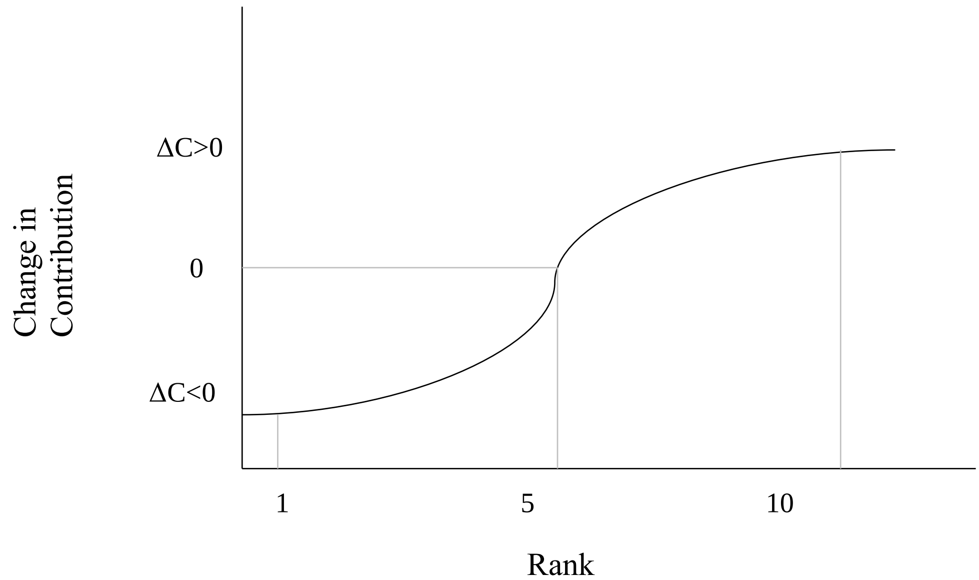


Figure 1.16. Schematic representation of rank effects in the C-experiment.

With this in mind, we used a novel approach for this analysis and constructed, for each participant, a graph that plots their change in contribution as a function of their rank (Figure 1.17, Figure 1.18). This graph captures each participant's response to their rank position within the feedback profile of the previous round. The graph consists of pairs of numbers: the x axis is the participant's ranked position in the feedback profile in any given round, and the y axis is the difference in contributions between the current round and the next round. Therefore, if the value on the y axis is positive it means that at that rank position the participant decided to increase their contribution. A y-axis value of zero implies that a participant did not change their contribution between rounds. Observing these figures, we can find some initial evidence that many of our participants displayed rank sensitivity.

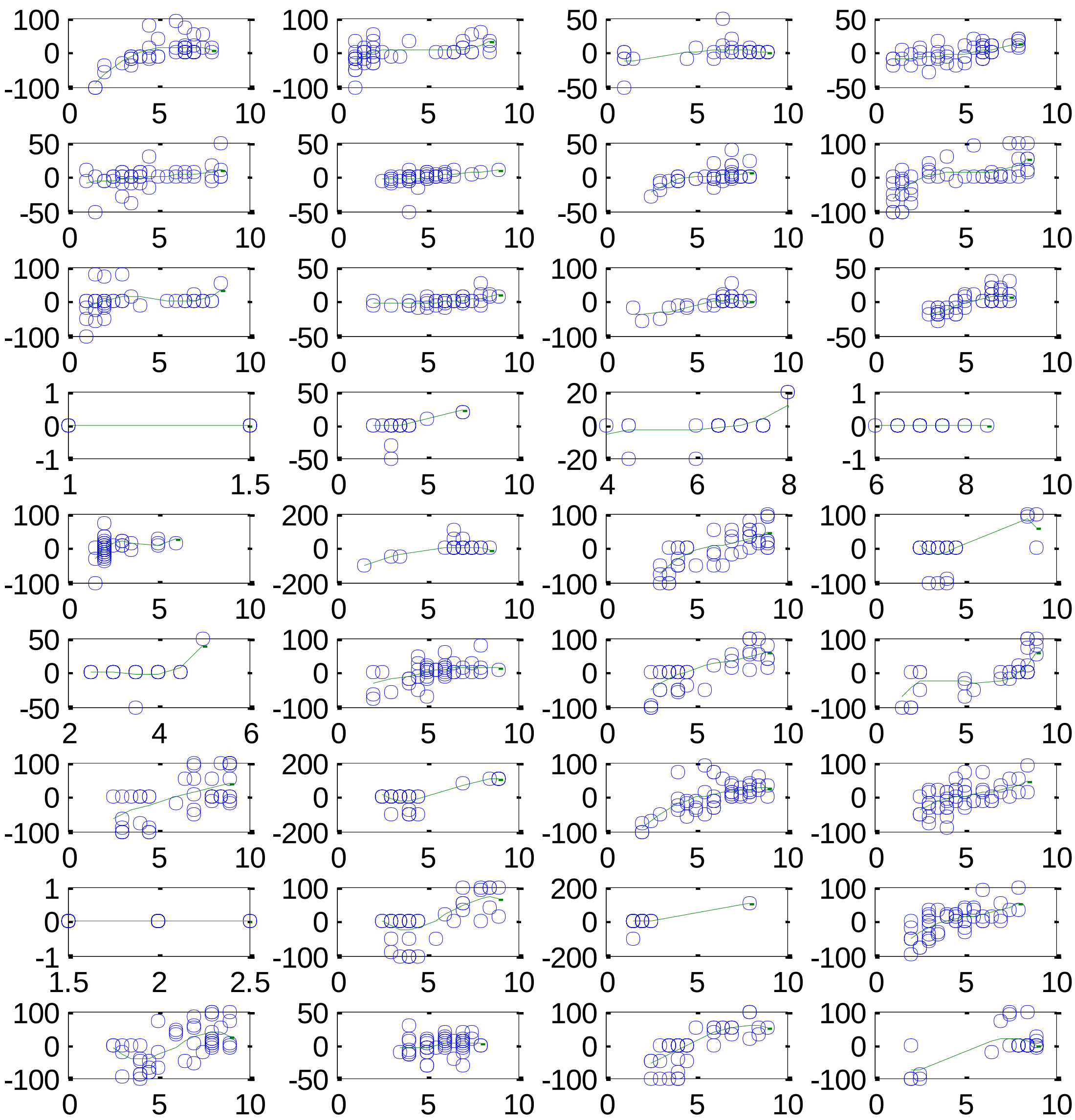


Figure 1.17. Changes in contribution as a function of rank. 36 plots of individual participants that completed 50 rounds of the game.

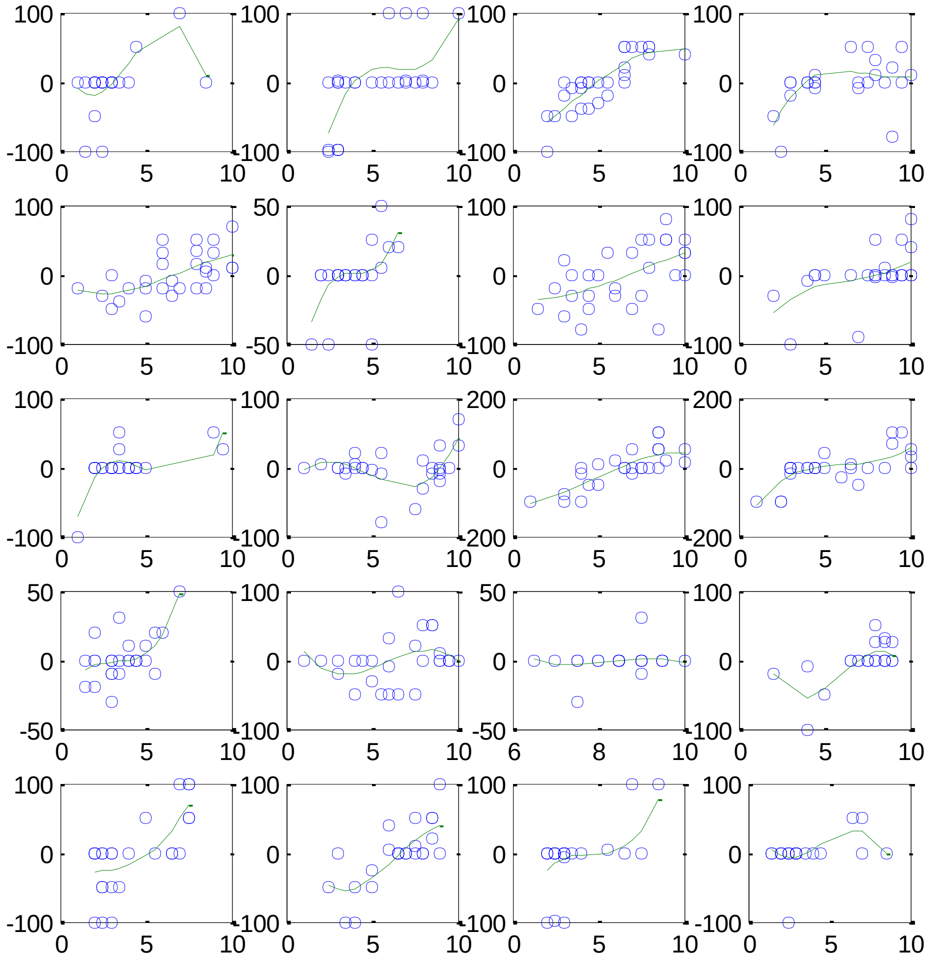


Figure 1.18. Changes in contribution as a function of rank. 20 plots of individual participants that completed 30 rounds of the game.

Setting the fitted 3rd order polynomial to zero and solving for x the roots of the polynomial can identify the positions of stable ranks for each individual from Figure 1.18. Here, the rank positions between 1 and 10 correspond to the highest and lowest contribution respectively. If multiple participants contributed the same amount then the rank position was given as the mean of the rank positions that the participants occupied.

Figure 1.19 shows the histogram of stable rank preferences. It shows a large peak around rank 5. This suggests that the most common stable preference was to be in the middle of the feedback. A rank of 5 does not correspond to 50 points which is the middle of the range (0-100) of possible contributions but to the middle contribution at a given round. Assuming participants' internal preferences were constant, then this leads to the identification of a stable ranked position i.e., a rank at which a participant had no incentive to change his or her contribution. However, one can conduct a similar type of analysis and show a similar pattern between rounds whereby the change in contribution is affected by the mean of contributions in the previous round.

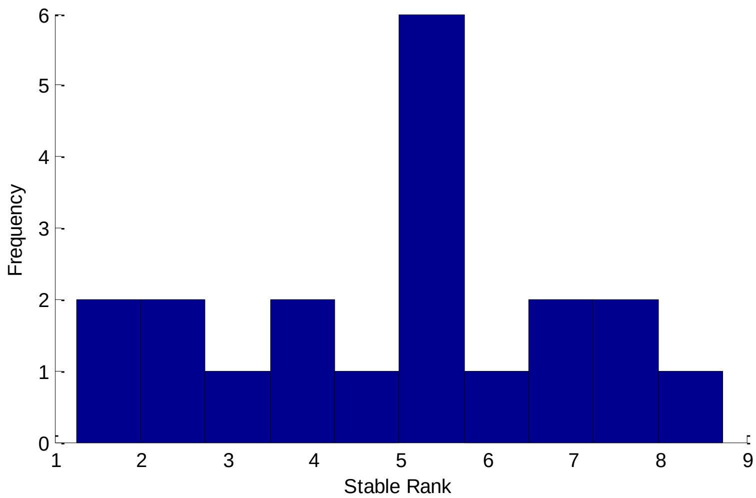


Figure 1.19. Histogram of stable ranks.

To distinguish between the effects of mean and rank we also ran an aggregate random effects regression with one's own contribution in the previous round, average of others in the previous round and rank cubed in the previous round as independent variables, trial as a random effect and contribution in the current round as dependent variable. The regression took the form

$$C_{ij} = \beta_0 + \beta_1\, Own\ C_{j-1} + \beta_2 Average\ of\ others_{j-1} + \beta_3 Own\ Rank^3{}_{j-1} + \varepsilon_i + \varepsilon_{ij,} \qquad (15)$$

where C_{ij} is the contribution of individual i for round j. The results of this regression are reported in Table 1.4.

Table 1.4. Aggregate random effects regression results for contributions in the current round depending on an individual's contribution in the previous round, the average contribution of others in the previous round, and an individual's relative rank postion cubed in the previous round.

	Estimate	Standard Error	t value	p value
(Intercept)	14.4	4.2	3.5	0.834
Own Contribution	0.33	0.04	8.25	<.001
Average Contribution	0.425	0.007	60.71	<.001
Rank (cubed)	-8.8	3.6	2.4	0.015
Obs:2344				

Own contributions in the previous affected significantly and positively contributions in the current round. Increasing the average contribution of others in the previous round also affected contributions in the current round significantly and positively. Rank cubed was found to have the strongest effect compared to rank but still the result is marginal (p = .015). The fact that rank cubed gave us the best result suggests that the observed marginal effect of rank came from ranks which were high. Participants who were at relative high rank positions in the previous rounds decreased their contributions more in the current round.

Rank here was defined from 0 to 1, corresponding to the lowest and highest contribution respectively. The rank coefficient being negative here supports the conclusion of the individual analysis suggesting that people tend to revert to a stable rank. If the rank is high in the previous round then it would have a negative effect in their contribution in the current round. Decreasing rank in the previous round, increases contribution in the current round ceteris paribus.

1.16 Discussion

Even though we tested bigger groups of players we found similar patterns of decreasing contributions in the C-experiment compared to studies that use only 4 players. With these bigger size groups, we also found a significant effect of the mean of the feedback on aggregate contributions, providing evidence for conditioning on the mean cooperative strategies. Nevertheless, no significant effects of standard deviation were found as in the P-experiment. Although this is not evidence against a rank-based model, an effect of standard deviation might not have been captured because the range of standard deviation in the experiment was small (it ranged approximately between 36 and 45 among all rounds). The use of confederates could be useful in that regard, allowing for a bigger range of standard deviation of contributions. However, this was predominately achieved in the P-experiment where we controlled for different aspects of the contribution profiles and thus the purpose of the C-experiment was to see what happens “in the wild”.

Although from Figure 1.17 and Figure 1.18 we can observe a lot of our participants being fit well by a sigmoidal curve running a mixed model showed marginal effects for rank. Thus, our results remain mixed and inconclusive. Moreover, it was clear that the strong effect of trial number precluded the possibility of treating the C-experiment as a series of one-shot games, not allowing us to conduct or test the social norms models presented in the P-experiment. Additionally, we could not rule out possible confounding effects due to strategic considerations or beliefs in this game. Due to these, we focused our analysis on the P-experiment presented in the previous section in which temporal effects were controlled for by the random ordering.

We believe that it is difficult to isolate the effects of the mean and rank given that participants were presented with overlapping information; participants saw feedback regarding what others did in the previous round, the average of all contributions in the previous round, their earnings in the previous round, and all the contributions of all group members including themselves in descending order.

Future studies could compare between treatments, where one treatment receives feedback about others’ contributions in the previous round without giving information about the mean, one treatment where information about the mean in the previous round is presented as well as all the individual contributions (but participants cannot see themselves in the distribution and a third treatment where participants would see the full list of contributions of others and their explicit (in numerical terms and not just visually) rank position within that feedback, testing specifically for rank effects and leaving every other information aside.

1.17 References

Aldrovandi, S., Brown, G.D.A., & Wood, A.M. (2015). Social norms and rank-based nudging: Changing willingness to pay for healthy food. *Journal of Experimental Psychology: Applied, 21,* 242-252.

Aldrovandi, S., Wood, A., Maltby, J., & Brown, G. D. A. (2015). Students' concern about indebtedness: A rank based social norms account. *Studies in Higher Education*, *40,* 1307-1327.

Allcott H. (2011). Social norms and energy conservation. *Journal of Public Economics 95,* 1082-1095.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104*(10), 3003-3037.

Alpizar, F., Carlsson, F., Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics 92*, 1047-1060.

Amir, O., Rand, D. G., & Gal, Y. A. K. (2012). Economic games on the Internet: The effect of $1 stakes. *PLoS ONE*, *7,* 31461.

Anderson, L., Mellor, J., & Milyo, J. (2004). *Do Liberals Play Nice? The Effects of Party and Political Ideology in Public Goods and Trust Games* (No. 07). Department of Economics, College of William and Mary.

Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics*, *37*(3), 291-304.

Arechar, A.A., Gächter, S., Molleman, L., 2018. Conducting interactive experiments online. *Experimental Economics*, *21*(1), 99-131.

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law Economics & Organization*, *29*(5), 992-1022.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, *28*(8), 1041-1055.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.

Bigoni, M., Suetens, S. (2012). Feedback and dynamics in public good experiments. *Journal of Economic Behavior & Organization 82*, 86–95.

Bilderbeck, A. C., Brown, G. D.A., Read, J., Woolrich, M., Cowen, P. J., Behrens, T. E., & Rogers, R. D. (2014). Serotonin and social norms: Tryptophan depletion impairs social comparison and leads to resource depletion in a multiplayer harvesting game. *Psychological Science*, *25*(7), 1303-1313.

BIT staff. (2013, November 22). *Behavioural Insights Tax Trials Win Civil Service Award.* Retrieved from http://www.behaviouralinsights.co.uk/tax/behavioural-insights-tax-trials-win-civil-service-award/.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166-193.

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471-475.

Boone, C., De Brabander, B., & Van Witteloostuijn, A. (1999). The impact of personality on behavior in five Prisoner's Dilemma games. *Journal of Economic Psychology*, *20*(3), 343-377.

Briesch, R. A., Krishnamurthi, L., Mazumdar, T., & Raj, S. P. (1997). A comparative analysis of reference price models. *Journal of Consumer Research, 24*(2), 202-214.

Brown, G. D.A., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does wage rank affect employees' wellbeing? *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355-389.

Brown, G.D.A., & Matthews, W.J. (2011). Decision by sampling and memory distinctiveness: Range effects from rank-based models of judgment and choice. *Frontiers in Psychology, 2,* 299.

Brown, G.D.A., Wood, A.M., Ogden, R.S., & Maltby, J. (2015). Do student evaluations of university reflect inaccurate beliefs or actual experience? A relative rank model. *Journal of Behavioral Decision Making, 28,* 14–26.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3-21.

Campbell, W. K., Bush, C. P., Brunell, A. B., & Shelton, J. (2005). Understanding the social costs of narcissism: The case of the tragedy of the commons. *Personality and Social Psychology Bulletin*, *31*(10), 1358-1368.

Carpenter J. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior 60,* 31-51.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics, 14*(1), 47-83.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review, 100*(4), 1358-1398.

Cheung, S. L. (2014). New insights into conditional cooperation and punishment from a strategy method experiment. *Experimental Economics*, *17*(1), 129-153.

Clark, A. E., & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics, 61*(3), 359-381.

Croson, R. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry, 45*(2), 199–216.

Croson, R., and J. Shang. (2008). The impact of downward social information on contribution decisions. *Experimental Economics, 11*(3), 221-233.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169-193.

Diederich, J., Goeschl, T., & Waichman, I. (2016). Group size and the (in) efficiency of pure public good provision. *European Economic Review, 85,* 272-287.

Dufwenberg, M., Gächter, S., Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior, 73*, 459-478.

Everett, J. A. (2013). The 12 item social and economic conservatism scale (SECS). *PloS one*, *8*(12), e82131.

Fallucchi, F., Luccasen, R. A., & Turocy, T. (2017). Behavioural types in public goods games: A re-analysis by hierarchical clutering, (No. 17-01R). *School of Economics, University of East Anglia*, Norwich, UK.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*(3), 817-868.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 1*, 117-140.

Fiala, L., & Suetens, S. (2017). Transparency and cooperation in repeated dilemma games: a meta study. *Experimental Economics, 20*(4), 755-771.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397-404.

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review, 100*(1), 541-556.

Gächter, S., Nosenzo, D., Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association, 11*, 548-573.

Gächter, S., Kölle, F., Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour, 1*(9), 650.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., & Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet*, *387*(10029), 1743-1752.

Hardin, G. (1968). The Tragedy of the Commons. *Science, 162*(3859), 1243-1248.

Hartig, B., Irlenbusch, B., & Kölle, F. (2015). Conditioning on what? Heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics, 55*, 48-64.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *The American Journal of Psychology*, *60*(1), 1-29.

Helson, H. (1964a). *Adaptation-level theory*. New York: Harper & Row.

Helson, H. (1964b). Current trends and issues in adaptation-level theory. *American Psychologist*, *19*, 26–38.

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598-603.

Hopkins, E., & Kornienko, T. (2004). Running to keep in the same place: Consumer choice as a game of status. *American Economic Review, 94*(4), 1085-1107.

Hyman, H. (1942). The psychology of subjective status. *Psychological Bulletin, 39*, 473-474.

Isaac, M.R., Walker, J.M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, *103*, 179-199.

Isaac, M.R., Walker, J.M., Williams, A.W. (1994). Group-size and the voluntary provision of public goods - experimental evidence utilizing large groups. *Journal of Public Economics, 54*, 1-36.

Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *The Scandinavian Journal of Economics*, *102*(1), 23-39.

Koole, S. L., Jager, W., van den Berg, A. E., Vlek, C. A., & Hofstee, W. K. (2001). On the social nature of personality: Effects of extraversion, agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin*, *27*(3), 289-301.

Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*, 495-524.

Ledyard, J. O. (1994). *Public Goods: A Survey of Experimental Research* (No. 9405003). EconWPA.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Sage Publications.

López-Pérez, R., 2008. Aversion to norm-breaking: A model. *Games and Economic Behavior, 64*, 237-267.

Lu, L., & Argyle, M. (1991). Happiness and cooperation. *Personality and Individual Differences*, *12*(10), 1019-1030.

Mailath, G. J., & Samuelson, L. (2006). *Repeated games and reputations: long-run relationships*. Oxford university press.

Maltby, J., Wood, A. M., Vlaev, I., Taylor, M. J., & Brown, G. D. A. (2012). Contextual effects on the perceived health benefits of exercise: The exercise rank hypothesis. *Journal of Sport and Exercise Psychology, 34*(6), 828-841.

Martin, R., & Randal, J. (2008). How is donation behaviour affected by the donations of others? *Journal of Economic Behavior & Organization 67*, 228-238.

Martin, R., & Randal, J. (2010). How social norms, price and scrutiny influence donation behaviour: evidence from four natural field experiments. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The science of giving: Experimental approaches to the study of charity*. New York: Taylor & Francis.

Mazumdar, T., Raj, S. P., & Sinha, I. (2005). Reference price research: review and propositions. *Journal of Marketing*, *69*(4), 84-102.

Mellers, B. A. (1982). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General*, *111*(2), 242–270.

Mellers, B. A. (1986). "Fair" allocations of salaries and taxes. *Journal of Experimental Psychology: Human Perception and Performance, 12*(1), 80-91.

Melrose, K. L., Brown, G. D. A., & Wood, A. M. (2012). Am I abnormal? Relative rank and social norm effects in judgments of anxiety and depression symptom severity. *Journal of Behavioral Decision Making*, *26*(2), 174-184.

Messick, D. M., Wilke, H., Brewer, M. B., Kramer, R. M., Zemke, P. E., & Lui, L. (1983). Individual adaptations and structural change as solutions to social dilemmas. *Journal of Personality and Social Psychology*, *44*(2), 294.

Mestelman, S., & Feeny, D. (1988). Does ideology matter?: Anecdotal experimental evidence on the voluntary provision of public goods. *Public Choice*, *57*(3), 281-286.

Neighbors, C., Larimer, M. E., & Lewis, M. A. (2004). Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. *Journal of Consulting and Clinical Psychology*, *72*(3), 434.

Niedrich, R. W., Sharma, S., & Wedell, D. H. (2001). Reference price and price perceptions: A comparison of alternative models. *Journal of Consumer Research, 28*(3), 339–354.

Niedrich, R. W., Weathers, D., Hill, R. C., & Bell, D. R. (2009). Specifying price judgments with range-frequency theory in models of brand choice. *Journal of Marketing Research, 46*(5), 693–702.

Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior, 68*(2), 689–702.

Olson, M. (1965). *Logic of collective action: Public goods and the theory of groups.* (Harvard economic studies. v. 124). Harvard University Press.

Olver, J. M., & Mooradian, T. A. (2003). Personality traits and personal values: A conceptual and empirical integration. *Personality and Individual Differences*, *35*(1), 109-125.

Panchanathan, K., Frankenhuis, W. E., & Silk, J. B. (2013). The bystander effect in an N-person dictator game. *Organizational Behavior and Human Decision Processes, 120*(2), 285-297.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411-419.

Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. (1960). Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *Journal of Experimental Psychology, 60*(2), 65–77.

Parducci, A. (1965). Category judgment: a range-frequency model. *Psychological Review*, *72*(6), 407-418.

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology, 89(*2*),* 427–452.

Parks, C. D., Rumble, A. C., & Posey, D. C. (2002). The effects of envy on reciprocation in a social dilemma. *Personality and Social Psychology Bulletin*, *28*(4), 509-520.

Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode". *The Annals of Mathematical Statistics 33*(3), 1065.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan press.

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics, 26*(3), 171-176.

Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function". *The Annals of Mathematical Statistics 27*(3), 832.

Samuelson, P. A. (1954). The Pure Theory of Public Expenditure. *The Review of Economics and Statistics, 36*(4), 387-389.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429-434.

Seithe, M., Morina, J., & Glöckner, A. (2016). Bonn eXperimental System (BoXS): An open-source platform for interactive experiments in psychology and economics. *Behavior Research Methods*, *48*(4), 1454-1475.

Sell, J., and R. K. Wilson. (1991). Levels of information and contributions to public goods. *Social Forces, 70*(1), 107-124.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments, S. *Beiträge zur experimentellen Wirtschafts-Forschung. Tübingen: JCB Mohr*, 136–168.

Shang, J., and R. Croson. (2009). A field experiment in charitable contribution: the impact of social information on the voluntary provision of public goods. *Economic Journal, 119*, 1422-1439.

Smith, R. H., Diener, E., & Wedell, D. H. (1989). Intrapersonal and social comparison determinants of happiness: A range-frequency analysis. *Journal of Personality and Social Psychology, 56*(3), 317–325.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by Sampling. *Cognitive Psychology, 53*(1), 1-26.

Taylor, M. J., Vlaev, I., Maltby, J., Brown, G. D. A., & Wood, A. M. (in press). Improving social norms interventions: Rank-framing increases excessive alcohol drinkers' information-seeking. *Health Psychology*.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Van Den Berg, P., Molleman, L., Junikka, J., Puurtinen, M., & Weissing, F. J. (2015). Human cooperation in groups: variation begets variation. *Nature: Scientific Reports, 5*, 16144.

Volk, S., Thöni, C., & Ruigrok, W. (2011). Personality, personal values and cooperation preferences in public goods games: A longitudinal study. *Personality and individual Differences, 50*(6), 810-815.

Weimann J. (1994). Individual behavior in a free riding experiment. *Journal of Public Economics 54*(2), 185-200.

Wilson, R. K., and J. Sell. (1997). “Liar, liar...”: cheap talk and reputation in repeated public goods settings. *Journal of Conflict Resolution, 41*(5), 695-717.

Wood, A. M., Brown, G. D. A., & Maltby, J. (2011). Thanks, but I'm used to better: A relative rank model of gratitude. *Emotion, 11*(1), 175.

Wood, A. M., Boyce, C. J., Moore, S. C., & Brown, G. D. A. (2012). An evolutionary based social rank explanation of why low income predicts mental distress: A 17 year cohort study of 30,000 people. *Journal of Affective Disorders, 136*(3), 882-888.

Wood, A. M., Brown, G. D. A., & Maltby, J. (2012). Social norm influences on evaluations of the risks associated with alcohol consumption: Applying the rank-based decision by sampling model to health judgments. *Alcohol and Alcoholism, 47*(1), 57-62.

1.18 Appendices

1.18.1 Appendix I

Instructions P-experiment

Thank you for agreeing to help us with our research on decision making.

In this study, you will take part in a decision-making task. Instructions will be provided on exactly how to do this shortly. After that, you will also be asked to provide to fill in some short questionnaires.

The aim of our research is to investigate decision making.

The study will take approximately 30 minutes to complete.

If at any point you wish to withdraw you can do so although payment will only be given to those who complete the entire experiment. Your payment is in two parts. The first part is a constant fee and the second part is a bonus payment that will be calculated based on your responses and will be explained later.

Questions or concerns

If you have any questions, please do not hesitate to contact the researchers at the email address provided, and we will assist you to the best of our abilities.

Enquiries to:
e.achtypi@warwick.ac.uk

During the experiment you will not be asked to reveal your identity and all data obtained from your decisions will be kept anonymous. All data will be handled by experimenters only, and will not be shared with any third party.

Before proceeding with the experiment, you:

□ Confirm that you have read and understood the information above.

□ Agree to take part in this study, and will not hold the Department of Psychology at the University of Warwick responsible for any mishap.

By selecting "Yes" at the bottom of this consent form, you agree that have read and understood the information presented above, and give full consent to proceed with the study.

Please do not talk to anyone else while completing this study. We are interested in your individual opinions.

Please read the instructions carefully. After the instructions there will be 2 examples. Following the examples, you will be asked to answer a short quiz to check your understanding of the decision situation. You will not be able to proceed to the HIT until you answer these questions correctly.

In order to confirm you have read these instructions, rather than clicking the "Next" button, first click the "A" on the University of Warwick logo above which will highlight green. Only once this has been completed you will be able to continue.

Once again, thank you for taking part in this study; your participation is greatly appreciated.

Decision Situation

Please read the following description of the Decision Situation carefully, as it is an essential part of the task you will be required to complete.

You will learn how the task will be conducted later. Firstly, we would like to introduce you to the decision situation.

The decision situation involves a group of ten people. Each member of the group is given 100 points and must decide what to do with their points.

You and the other group members must each state simultaneously, independently and anonymously how many of the 100 points they would like to contribute into a Group Pot without seeing the choices of anybody else. You keep any points that you do not contribute to the Group Pot.

Each point you and the other members put into the Group Pot is multiplied by five. The resulting total points are then split evenly among all group members.

Your Total Payout is comprised of two parts: the points you kept and your share of points from the Group Pot. The Total Payout of all the other players is calculated in exactly the same way as yours. We will illustrate this with a few examples on the next page.

In order to confirm you have read the Decision Situation, rather than clicking the "Next" button, first click the "De" on the Decision Situation title above which will highlight green. Only once this has been completed you will be able to continue.

Here are some examples to illustrate how the decision situation works:

<u>Example 1:</u>

You contribute 20 points to the group pot and keep 80. The other players contribute the following amounts.

	Group pot contribution
Player A	60
Player B	56
Player C	51
Player D	45
Player E	41
Player F	37
Player G	32
Player H	30
Player I	28

You	20		

Sum of contributions to the group pot	400		
Payout from group pot	400	x 5 =	2000
Each player's payout from the group pot	2000	/ 10 =	200

Payout

The total contributed to the group pot is 400. This comes from total contribution of the other 9 people contributing a total of 380 and your contribution of 20 (380+20=400). This is multiplied by 5 to give 2000. Each member of the group therefore receives a payout from the pot of 2000/10 = 200.

You receive a payout of 200 points from the group pot. You also keep the 100-20=80 points you did not contribute giving you a total payout of 80+200=280 points.

The other players also receive a payout of 200 from the group pot. Their total payouts also include the amounts they keep. So, the other players receive a total payout in the range 240 - 272 points.

<u>Example 2:</u>

You contribute 50 points to the group pot and keep 50. The other players contribute the following amounts.

	Group pot contribution
Player A	75
Player B	71
Player C	64
Player D	54
Player E	50
Player F	46
Player G	36
Player H	29
Player I	25
You	50

Sum of contributions to the group pot	500		
Payout from group pot	500	x 5 =	2500
Each player's payout from the group pot	2500	/ 10 =	250

Payout

The total contributed to the group pot is 500. This comes from total contribution of the other 9 people contributing a total of 450 and your contribution of 50 (450+50=500). This is multiplied by 5 to give 2500. Each member of the group therefore receives a payout from the pot of 2500/10 = 250.

You receive a payout of 250 points from the group pot. You also keep the 100-50=50 points you did not contribute giving you a total payout of 50+250=300 points.

The other players also receive a payout of 250 from the group pot. Their total payouts also include the amounts they keep. So, the other players receive a total payout in the range 275 - 325 points.

<u>Example 3:</u>

You contribute 80 points to the group pot and keep 20. The other players contribute the following amounts.

	Group pot contribution
Player A	74
Player B	71
Player C	67
Player D	65
Player E	60
Player F	55
Player G	53
Player H	49
Player I	46

You	80		

Sum of contributions to the group pot	620		
Payout from group pot	620	x 5 =	3100
Each player's payout from the group pot	3100	/ 10 =	310

Payout

The total contributed to the group pot is 620. This comes from total contribution of the other 9 people contributing a total of 540 and your contribution of 80 (540+80=620). This is multiplied by 5 to give 3100. Each member of the group therefore receives a payout from the pot of 3100/10 = 310.

You receive a payout of 310 points from the group pot. You also keep the 100-80=20 points you did not contribute giving you a total payout of 310+20=330 points.

The other players also receive a payout of 310 from the group pot. Their total payouts also include the amounts they keep. So, the other players receive a total payout in the range 336 - 364 points.

<u>Quiz</u>

In order to make sure that you have understood the Decision Situation we would like to ask you to complete the following quiz, consisting of three parts. You need to get each part right to move on to the next part. You can only proceed to the actual task of the HIT once you have correctly answered all questions.

Let us remind you of the basic rules of the game:

You are a member of a group consisting of ten people.

Each member of the group is given 100 points and must simultaneously and anonymously decide how many of their points to contribute to a group pot.

Each point you and the other members contribute to the group pot will be multiplied by five. The resulting points will be evenly distributed among all group members regardless of who contributed what.

You keep each point that you do not contribute to the group pot.

Your total points are comprised of two parts: your share from the group pot payout plus the points you chose not to contribute.

You contribute 0 points to the group pot, and keep 100 points. The 9 other players all do the same, contributing 0 points to the group pot and keeping 100.

What is your total payout?

What is the total payout of each of the other players?

<u>Quiz - Part 2</u>

Let us remind you of the basic rules of the game:

You are a member of a group consisting of ten people.

Each member of the group is given 100 points and must decide simultaneously and anonymously how many of their points to contribute to a group pot.

Each point you and the other members contribute to the group pot will be multiplied by five. The resulting points will be evenly distributed among all group members regardless of who contributed what.

You keep each point that you do not contribute to the group pot.

Your total points are comprised of two parts: your share from the group pot payout plus the points you chose not to contribute.

You contribute 100 points to the group pot, and keep 0 points. The 9 other players all do the same, contributing 100 points to the group pot and keeping 0.

What is your total payout?

What is the total payout of each of the other players?

Quiz - Part 3

Let us remind you of the basic rules of the game:

You are a member of a group consisting of ten people.

Each member of the group is given 100 points and must decide simultaneously and anonymously how many of their points to contribute to a group pot.

Each point you and the other members contribute to the group pot will be multiplied by five. The resulting points will be evenly distributed among all group members regardless of who contributed what.

You keep each point that you do not contribute to the group pot.

Your total points are comprised of two parts: your share from the group pot payout plus the points you chose not to contribute.

The 9 other players together contribute a total of 400 to the group pot. You contribute 20 to the group pot, keeping 80.

What is your total payout?

The 9 other players together contribute a total of 400 to the group pot. You contribute 60 to the group pot, keeping 40.

What is your total payout?

Introduction to the Task

Thank you for your attention so far. You've finished the learning part of the study as well as the quiz.

This part of the study is split into two tasks. In the first task you will be asked how much you would contribute to the Group Pot in the decision situation we have described to you.

In the second task you will be asked what you would choose to do in a variety of situations like the ones you have seen in the examples. You will be presented with some possible combinations of contributions. You will be asked to specify how much you want to contribute given these circumstances. You should think of these combinations as distinct scenarios, separate from each other.

Your responses to these two tasks will determine your bonus payment. One of your answers will be randomly selected and played out for real as per the rules of the decision situation. Your total payout for that situation will then be converted into a bonus payment at a rate of 100 points to $1.

If your response to task one is chosen, then the responses of the other group members used to determine your payment will be made up of real answers given by other people participating in this same study. If one of your responses to task two is chosen, then the contributions of the other group members will be the combination of contributions you saw when giving that response. Please note that all responses across both tasks are equally likely to be chosen.

Since you do not know which of your responses will be played out for real, you should treat all decisions in the HIT as if they determine your payment.

Here is an example:
The situation chosen to determine your bonus payment had a total contribution by the other group members of 700. Your choice of contribution to the Group Pot for this situation is 40. Your Total Payout is therefore 430. This is made up of 60 points that you kept and 370 points from the Group Pot. This is would then give you a bonus payment of $4.30 = $1*430 points/100 points.

We can learn a lot from the opinions you give us in this part of the study, so please consider each answer carefully.

Part One of the HIT

You are taking part in the decision situation and you are given 100 points.

How many of your 100 points would you contribute to the Group Pot?

Part Two of the HIT (Presented on a separate screen)

You will now be presented with some sets of possible contributions in descending order. In each case you must decide how much you would like to contribute to the Group Pot. Each set of contributions is not connected to the others and you should think of them as distinct scenarios.

Suppose the contributions of the other group members are the following:

	Group Pot Contribution
Player A	96
Player B	89
Player C	76
Player D	74
Player E	70
Player F	66
Player G	64
Player H	51
Player I	44

How many of the 100 points do you contribute to the group pot?

Suppose the contributions of the other group members are the following

	Group Pot Contribution
Player A	77
Player B	70
Player C	57
Player D	55
Player E	51
Player F	47
Player G	45
Player H	32
Player I	25

How many of the 100 points do you contribute to the group pot?

/////Participants saw next the rest 25 contribution profiles in random order and separate screens. //////

Thank you for completing the task. We would now like to ask you some survey questions related to the task you just completed. Following that you will complete some short questionnaires. All data will be kept strictly anonymous and private.

Your answers are very valuable to us.

Thank you for your participation, we are grateful for your contribution to our study. Your unique completion code is:

${e://Field/code}

Please enter this code into Amazon MTurk to complete the HIT.

1.18.2 Appendix II

Contribution Profiles

Table T 1.2 shows the profiles used in the experiment, along with some basic statistics. They are described in the form $\Sigma(\mu, \sigma, s)$, where μ, σ and s are the mean, standard deviation and skewness. We describe the profiles with these three statistics because they are formally independent of each other though other aspects, such as the range, were taken into account when designing them, detailed below. Our profiles were designed to cover a large range of possibilities. As a consequence, and according to the strategy method paradigm, many of the constructed stimuli may only be observed rarely in genuine public goods games. To ensure that the profiles we constructed fairly represented the possibilities, we sought that the average of the means was approximately 50, $\bar{\mu} \approx 50$ (actual value 50.06) and that the average of the skewness was approximately zero,$\bar{s} \approx 0$ (-0.007). Finally, we note also that the average kurtosis was also approximately zero (-0.0916). To ensure that there was no confounding effect between these variables, we constructed them such that all correlations between them were zero, shown in Table T 1.1.

Table T 1.1. Correlations between contribution profile statistics. None are significant.

	μ	σ	S	K
μ	.	0.003	-0.08	-0.0009
σ	.	.	-0.007	-0.09
s	.	.	.	0.002

Many of the profiles with zero skew were constructed according to a symmetric principle. A profile $\Sigma(\mu, \sigma, s) = (d_1, \ldots, d_9)$, has eight differences of consecutive values, $\Delta_i = d_{i+1} - d_i$, the symmetric profiles were created by choosing $\Delta_1 = \Delta_8$, $\Delta_2 = \Delta_7$ and so on, which guaranteed zero skew. They did not contain any repeated values within. This ensured that rank-based models could differentiate between all rank positions in the profile, as well as avoiding any anchoring effects coming from repeated values. As well as satisfying these overall constraints, many of the profiles were constructed as series which controlled for other aspects of the distribution. For example, the pairs S1, S3, S5 and S2, S4, S6 correspond to unimodal and bimodal distributions respectively, in which we have controlled for the mean, skewness, median and range and varied the modality and standard deviation. The series S21 through S23 also gives a progression from unimodal to bimodal distributions, with a high range. The pairs S7, S9, S11 and S8, S10, S12 correspond to high and low skew distributions respectively, in which we have controlled for the mean, standard deviation and range, but varied the skewness. The pair S19 and S20 correspond to controlling all aspects of the distribution except for the mean, which is varied. The groups S24-27 correspond to distributions that have the same mean, skewness, and modality but vary in the standard deviation. The pairs S13-14 and S15-16 correspond to varying the standard deviation while keeping the mean low and high respectively. Finally, the pair S17 and S18 correspond to controlling the standard deviation and range but changing the skewness and the mean.

Table T 1.2. Contribution profiles used in the experiment.

Profiles	Mean	Std. Dev.	Skewness	Values								
S1	70	16.5	0	96	89	76	74	70	66	64	51	44
S2	70	22.4	0	96	93	91	89	70	51	49	47	44
S3	51	16.5	0	77	70	57	55	51	47	45	32	25
S4	51	22.4	0	77	74	72	70	51	32	30	28	25
S5	29	16.5	0	55	48	35	33	29	25	23	10	3
S6	29	22.4	0	55	52	50	48	29	10	8	6	3
S7	60	17.0	-1.53	73	72	71	69	66	63	60	45	21
S8	60	17.0	1.53	99	75	60	57	54	51	49	48	47
S9	51	17.0	-1.53	64	63	62	60	57	54	51	36	12
S10	51	17.0	1.53	90	66	51	48	45	42	40	39	38
S11	39	17.0	-1.53	52	51	50	48	45	42	39	24	0
S12	39	17.0	1.53	78	54	39	36	33	30	28	27	26
S13	20	7.0	0	30	28	25	23	20	17	15	12	10
S14	20	15.8	0	40	38	35	27	20	13	5	2	0
S15	80	7.0	0	90	88	85	83	80	77	75	72	70
S16	80	15.8	0	100	98	95	87	80	73	65	62	60
S17	33	34.4	0.92	98	76	50	30	16	10	8	5	4
S18	67	34.4	-0.92	96	95	92	90	84	70	50	24	2
S19	13	4.9	0	20	19	16	14	13	12	10	7	6

S20	87	4.9	0	94	93	90	88	87	86	84	81	80
S21	50.3	25.7	-0.04	97	64	62	59	53	41	38	36	3
S22	51.1	31.4	-0.11	97	78	74	68	50	32	26	22	3
S23	50	39.5	0	97	92	86	81	50	19	14	8	3
S24	50	4.9	0	57	56	53	51	50	49	47	44	43
S25	50.1	14.6	-0.02	71	68	59	53	51	47	41	32	29
S26	50.2	24.4	-0.03	85	80	65	55	52	45	35	20	15
S27	49.8	34.1	0.02	99	92	71	57	48	43	29	8	1

1.18.3 Appendix III

Individual Differences in the P-experiment

After the completion of the P-experiment participants were asked to complete some psychological scales relating to personality and conservatism. Detailed results of the analysis between contributions and these measurements can be found here.

Overview

Here we focus on the effects of individual differences to gain a more holistic understanding of the determinants of behaviour in the public goods P-experiment.

Our P-experiment provided an ideal platform to not only test the relationship between absolute contribution levels and individual measurements but also look at the people who were better fit by the rank model and those who were better fit by the mean relative model and try to find if they differed in other respects. More specifically, we focused on the effects of personality and conservatism.

Personality

Personality traits can be considered innate characteristics describing the nature of an individual compared to values that can be learned from the environment and others' behaviours and the interplay between nature and nurture (Olver & Mooradian, 2003).

Personality influences have been investigated in social dilemmas, for example, cooperation is higher for individuals low in narcissism (Campbell, Bush, & Brunell, 2005), envy (Parks, Rumble, & Posey, 2002) and high in sensation seeking and self-monitoring (Boone, Brabander, & van Witteloostuijn, 1999). Participants with high honesty–humility exhibited more cooperation in a prisoner's dilemma; participants low in honesty–humility free rode more. (Hilbig, Zettler, Leist, & Heydasch, 2013).

In Koole, Jager, Van den Berg, Vlek and Hofstee (2001) participants low in extraversion and participants high in agreeableness harvested less from a common resource than did those with high extraversion and low agreeableness. Those low in extraversion and those high in agreeableness decreased their extractions when the common resource was more depleted. Extraversion relates to being more sociable, active and talkative and agreeableness relating to trusting and tolerating others.

The longitudinal study by Volk, Thöni and Ruigrok (2011) is the only other study we are aware of that examined cooperation preferences in a P-experiment public goods and personality. In a series of three public goods games that were conducted over the course of five months, results showed that agreeableness was a significant predictor of cooperation. Participants low on agreeableness were more likely to be free-riders compared to participants high on agreeableness who were more likely to be conditional cooperators.

After the completion of the P-experiment participants were asked to complete the psychological scales discussed below relating to personality.

We measured our participant's personality profiles with the ten-item personality TIPI scale (Ten Item Personality Inventory), which includes two items for each of the Big-Five personality dimensions. Participants rated the extent to which trait applied to them on a 7-point scale (strongly disagree to strongly agree) (Gosling, Rentfrow, & Swann, 2003).

I see myself as:

1.Extraverted, enthusiastic.

2.Critical, quarrelsome.

3.Dependable, self-disciplined.

4.Anxious, easily upset.

5.Open to new experiences, complex.

6.Reserved, quiet.

7. Sympathetic, warm.

8.Disorganized, careless.

9.Calm, emotionally stable.

10.Conventional, uncreative.

(R denotes reverse-scored items):

Extraversion is measured by 1, 6R
Agreeableness by 2R, 7
Conscientiousness by 3, 8R
Emotional Stability by 4R, 9
Openness to Experiences by 5, 10R

Results Personality

The first analysis was an OLS regression with as its dependent variable the contributions of the individuals and as independent variable their personality measures. We incorporated the different contribution profiles participants saw as a random effect. The results of this regression are shown in the table below.

Table T 1.3. Aggregate random effects regression results for contributions depending on the Big Five personality items.

	Estimate	Standard Error	t statistic	p value
(Intercept)	34.80	3.13	11.10	<.001
Extraversion	-0.85	0.21	-4.06	<.001
Agreeableness	1.57	0.31	4.99	<.001
Openness	-0.23	0.26	-0.86	0.39
Conscientiousness	2.31	0.29	7.97	<.001
Emotional Stability	-0.43	0.26	-1.67	0.10
Obs: 8100				

These results demonstrate that all but two of the personality measures (openness, emotional stability) showed significant effects. More importantly, the effects of agreeableness and extraversion were in the same direction as others have found (Lu & Argyle, 1991). We also found a positive effect of conscientiousness (being organized, careful and taking tasks seriously). Running a regression with robust standard errors showed no differences in estimates nor standard errors and exhibited the same significant results.

Moving beyond this, we were interested in the difference between individuals whom we identified as mean relative winners and rank winners. To investigate this difference, we ran a logistic regression. Because we have a large sample size (of order one hundred in each class) we used all the personality measures. The model showed no significance (*Chisq* = 0.44, *p* = 0.99), with odds ratios detailed in the table below. Reference category was mean relative winner.

Table T 1.4. Logistic regression results.

	λ	Standard Error	Wald Z	p value	Exp(λ)
Extraversion	0.01	0.09	0.11	0.91	1.01
Agreeableness	0.00	0.15	0.03	0.98	1
Conscientiousness	0.03	0.13	0.27	0.79	1.03
Openness	-0.06	0.12	-0.52	0.60	0.94
Emotional Stability	-0.03	0.12	-0.23	0.82	0.97
Obs: 195					

We conclude that there were no significant differences in personality measures between rank winners and mean relative winners.

Ideology-Conservatism

There is not much published research on the effect of ideology on cooperation in experimental games. Although national survey results show that Democrats and liberals are more likely to favor spending on public programs, there is no evidence that ideology or political preferences affect contributions in a public goods game. In Anderson, Mellor and Milyo (2005), after playing a C-experiment public goods game subjects were asked to indicate the political party to which they belonged, the political party that best represented their interests (including Democrat, Republican, other, and none), and to rate their political ideology leanings on a scale from zero to 10 (0- extreme conservative, 5- moderate, 10 extreme liberal). Because liberals may not have behaved more compassionately in the artificially egalitarian setting, inequality among subjects (manipulation of the show-up fee) was induced. No effect of political party or ideology on public goods contributions were found or interactions of those measures with the inequality treatment. The other work we are aware of connecting ideology and cooperation (Mestelman & Feeny, 1988) was done with subjects all of whom had a negative attitude towards free-riding (group of scientists attending the second day of a common-property resource management conference). In this repeated public goods game complete free riding did not emerge quickly but the proportion of free riders increased steadily.

In order to investigate the effects of conservatism on contribution decisions we used the 12 Item Social and Economic Conservatism Scale (SECS) (Everett, 2013). This scale is modern and validated, and measures conservatism in two dimensions: social and economic. The SECS measures conservatism, regardless of party affiliation.

Participants were asked to indicate the extent to which they felt positive or negative towards the issues reported below on a scale from 0 to 100 (with 100 corresponding to extreme conservatism).

Abortion (reverse scored) (Social Conservatism),

Limited government (Economic Conservatism),

Military and national security (Social Conservatism),

Religion (Social Conservatism),

Welfare benefits (reverse scored) (Economic Conservatism),

Gun ownership (Economic Conservatism),

Traditional marriage (Social Conservatism),

Traditional values (Social Conservatism),

Fiscal responsibility (Economic Conservatism),

Business (Economic Conservatism),

the family unit (Social Conservatism),

Patriotism (Social Conservatism).

Results Conservatism

The first analysis was an OLS regression with as its dependent variable the contributions of the individuals and as independent variable their social and economic conservatism. We incorporated the different contribution profiles they saw as a random effect; the results of this regression are reported in the table below. Note that conservatism is measured on a scale from 0-100 which contributes to the low values of the coefficients.

Table T 1.5. Aggregate random effects regression results for contributions depending on the two conservatism items.

	Estimate	Standard Error	t statistic	p value
(Intercept)	50.16	2.76	18.88	<.001
Economic Conservatism	-0.094	0.023	-4.154	<.001
Social Conservatism	0.084	0.018	4.711	<.001
Obs:8100				

It is interesting that both these coefficients are significant, with similar magnitudes, but in opposite directions. Note that while the coefficients seem small, 0.094*100 is approximately 10, so going from minimum to maximum conservatism accounts for a difference in contributions of around 10. We once again ran a regression with robust standard errors which yielded similar results. Participants contributed differently given these different natures of conservatism, with individuals high on social conservatism perhaps feeling a social duty to contribute more and those individuals high on economic conservatism contributing less and thus being more rational in economic terms and closer to free riders. However, these results might not be realistic as these two measures of conservatism tend to correlate.

Again, we looked at the differences between rank and mean relative winners with a logistic regression. This had a significant fit (*Chisq* = 9.78, *p* = 0.0075). The results are shown in the table below. Reference category was mean relative winner.

Table T 1.6. Logistic regression results.

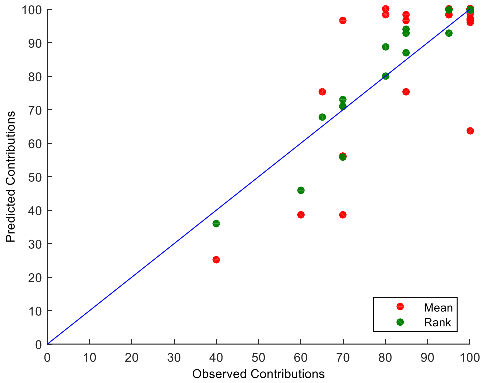
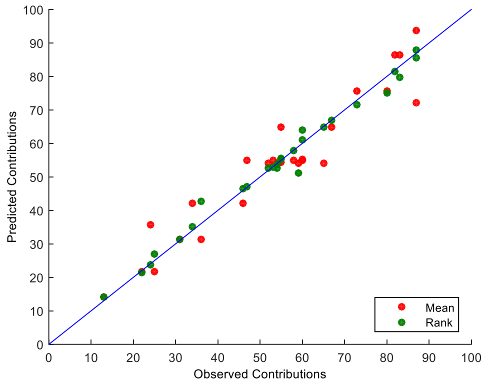
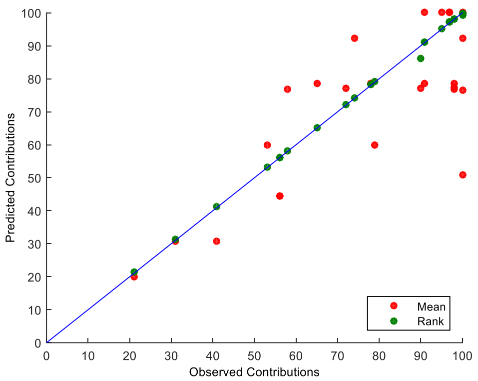
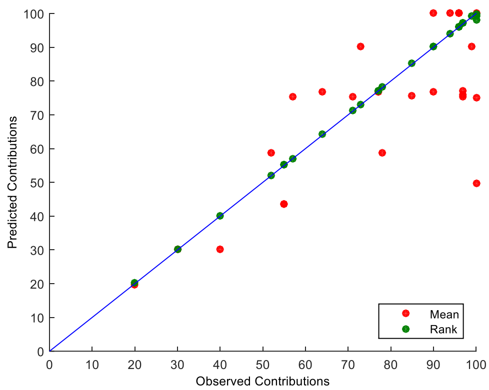
	λ	Standard Error	Wald Z	p value	Exp (λ)
Economic Conservatism	-0.035	0.011	-3.00	0.0027	0.97
Social Conservatism	0.018	0.008	2.10	0.035	1.02
Obs:195					

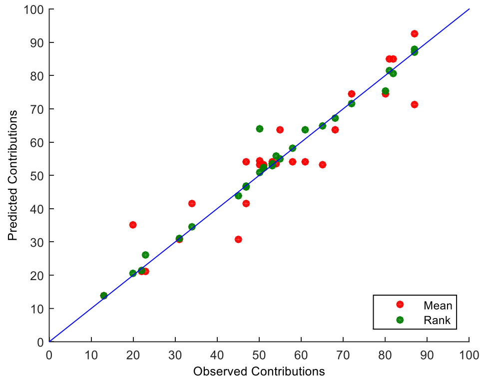
Again, the log ratios are small due to the scale of our conservatism measure. Of particular interest is that economic conservatism decreases the chance of being a rank winner, while social conservatism increases that chance but marginally. An increase of the economic conservatism score by one point makes it 0.97 times less likely that a participant is categorized as a rank winner instead of mean winner. Interestingly, we found that conservatism distinguishes between whether one was classified as a rank or mean-based individual.

1.18.4 Appendix IV

Supplementary figures for individual fits of the rank model and the mean relative model

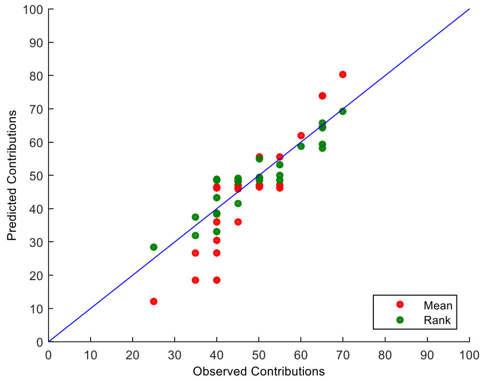
Table T 1.7. Comparison of predictions versus observations for rank and mean relative models for selected rank winners. Rank estimates and mean estimates in parentheses.

 Participant 8 (1, 1.93)	 Participant 33 (0.56, 1.08)
 Participant 135 (0.99, 1.54)	 Participant 148 (0.99, 1.51)



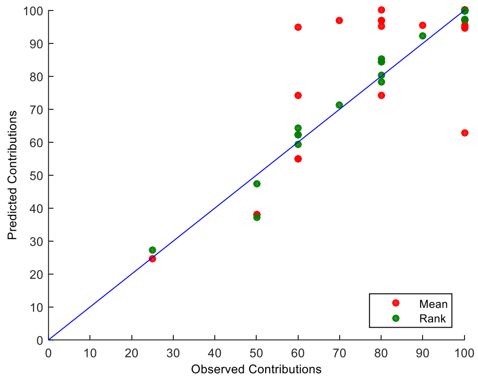
Participant 209

(0.56, 1.09)



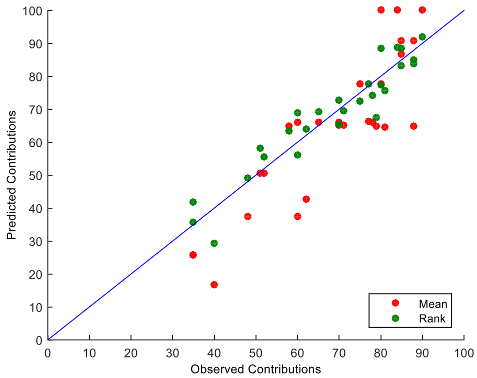
Participant 217

(0.48, 0.92)



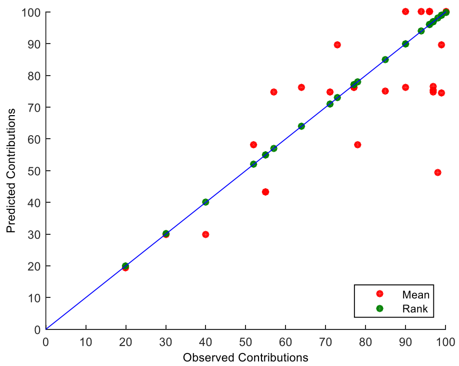
Participant 2

(1, 1.89)



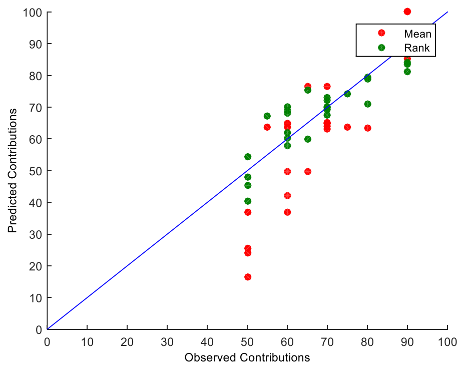
Participant 245

(0.79, 1.30)



Participant 256

(0.97, 1.49)



Participant 280

(0.74, 1.27)

1.18.5 Appendix V

Instructions for C-experiment

//////Introduction///////

Welcome to today's session. It will last approximately 1 hour.

We ask you to answer all sections as honestly and to the best of your ability.

If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore very important that you read these instructions carefully.

You are not allowed to communicate with other participants during the experiment.

If you have a question at any time, please raise your hand. A member of the experimenting team will come to you and answer it in private.

("Confirm & continue")

IMPORTANT

You are player ("…") in group ("…"). Please write down this information on the form provided.

/////INSTRUCTIONS////

You will be a member of a group of ten people. You will play a decision game consisting of 30 rounds {Group number was changed according to the number of participants that turned up in the lab as well as the rounds}.

All participants in the experiment will be divided in groups of ten members. Except for us - the experimenters - no one knows who is in which group.

At the start of each of the 30 rounds each group member is given 100 points. Both your and the other group members' task is to decide how many of the 100 points to contribute to a group project. You keep for yourself any points you do not contribute into the group project, and these points remain in your private account.

Each point that you and the other members contribute to the group project will be multiplied by five after everybody has decided how much to contribute. These multiplied points in the group project will be split evenly between all group members at the end of the round. It does not matter who has contributed points to the group project; everybody receives the same share from the project at the end of the round. Your contribution to the project therefore raises the income of the other group members and every other member's contribution raises your income. The more the group invests in the project, the greater the return to each member of the group.

Thus, the total number of points you will have at the end of each round is made up of two parts: the points you have kept in your private account, and your income from the group project. The points earned by each of the other players is calculated in exactly the same way as yours. Thus, each player's total points at the end of the round will be the number of points they kept in their private account, plus their share of the income from the group project.

Your total income in points = (100 - your contribution to the project) + 5/10*(total contributions to the project)

Feedback will be given after each round for the group you were in, showing how much other people in the group contributed to the group project, mean of contributions and your own points earned for that round.

Earnings

During the experiment your entire earnings will be calculated in points. At the end of the experiment the amount of points you have earned in a randomly selected round will be converted to sterling at the following rate: 1 point = 1 penny, so 100 points is one pound. Since you do not know which round will be chosen, you should act as if every round may be the one for which you are paid. This payment will constitute your entire earnings from this game and the £3 show up fee will be immediately paid to you in cash at the end of the session.

Next, we illustrate the above with a few examples

///////////EXAMPLES/////////////

Example 1:

Decision

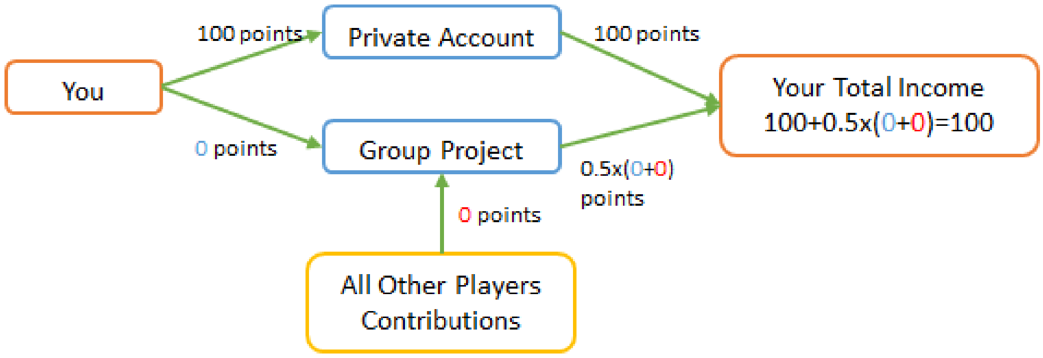
You keep 100 points in your private account and contribute zero to the group project. All the other players do the same.

Points earned

You and everyone else each keep 100 points from your private accounts and receive zero points from the group project as no player contributed to it.

This gives you 100 points from your private account and 0 points from the group project, which is a total of 100 points for you.

All other players get the same number of points as you do.



Example 2:

Decision

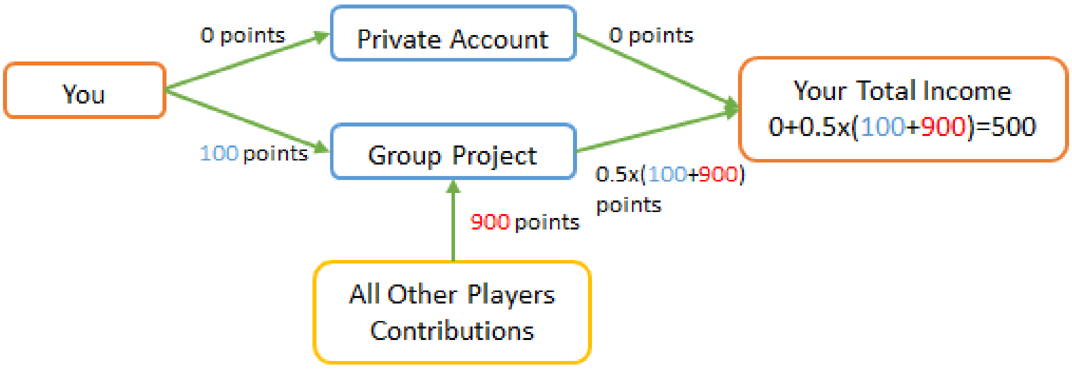
You keep zero points in your private account and contribute 100 points to the group project. All the other players do the same.

Points earned

The total amount contributed to the group project is the other nine players' contributions of 100 points each plus your contribution of 100 points to make a grand total of 9x100 + 100 = 1000 points. This is multiplied by 5 to give 5000. Each member of the group therefore receives an income from the project of 5000/10 = 500 points.

This gives you 0 points from your private account and 500 points from the group project, which is a total of 500 points for you.

All other players get the same number of points as you do.



Example 3:

Decision

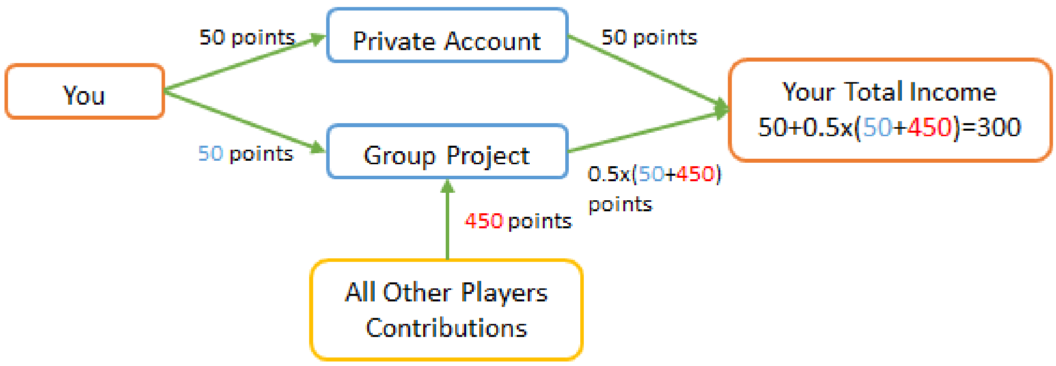
You and the other players each keep 50 points in your private accounts and contribute 50 points to the group project.

Points earned

The total amount contributed to the group project is the other nine players' contributions of 50 points each plus your contribution of 50 points to make a grand total of 9x50 + 50 = 500 points. This is multiplied by 5 to give 2500. Each member of the group therefore receives an income from the project of 2500/10 = 250.

This gives you 50 points from your private account and 250 points from the group project, which is a total of 300 points for you.

All other players get the same number of points as you do.



//////QUIZ/////

Quiz

Please answer the following questions. These serve to check your understanding of the decision situation and earnings calculations. When everyone has completed all the questions correctly, the experimental game will start.

Let us remind you of the basic rules of the game: You are a member of a group consisting of ten people. Each member of the group is given 100 points and must decide how many of their points to contribute to a group project. Each point that you do not invest will stay in your private account. Each point you and the other members invest in the group project will be multiplied by 5 and then shared equally between the members of the group. That is your total number of points is made up of two parts: points kept in your private account and your income from the group project.

Please input your answers. Only once all answers are correct, the button at the bottom changes from 'incorrect entry' to 'continue'.

If you contribute 0 points to the project and the other 9 participants each contribute 100 points to the project, what is the total number of points you earn?

{550,"correct entry"}

If you contribute 100 points to the project and the other 9 participants each contribute 0 points to the project, what is your total number of points you earn?

{50,"correct entry"}

Please press continue to start the experiment. You will be matched with 9 other people for 30 rounds.

At the beginning of each trial you will see an input-screen. You have to decide how many points you want to contribute to the project by typing a number between 0 and 100 in the input field. This field can be reached by clicking it with the mouse. As soon as you have decided how many points to contribute to the project, you have also decided how many points you keep for yourself: This is (100 - your contribution) points. After entering your contribution, you must press the Continue button with the mouse. Once you have done this your decision can no longer be revised.

After all members of your group have made their decision a following feedback screen will show you how much each member of the group has contributed to the group project. This screen will also show your total points earned (how many points you have earned in that trial) and the mean of contributions for that trial.

//////////Please press continue to start the experiment///////////

How much would you like to contribute from your 100 points to the group project? Please enter in the box below:

{"Continue", "Please wait for other players"}

///////// Inform the subjects about the results of each trial///////

////////Example of what participants saw after each round///////

//display("You are player "**…**". You contributed "**…**" Points. The average amount contributed is mean contribution "**…**" Points, your payoff for this round was ("**…**" Points.)

//display("The contributions from all players in decreasing order were:")

Player "E":70

Player "D":68

Player "H":48

Player "A":42

Player "G":30

Player "C":28

Player "F":25

Player "B":10

Player "I":5

Player "J":5

/////// Inform the participants about their final payoff///////

//display("Round "**…**" was randomly chosen to be played out. You contributed "**…**" Points. The other members contributed "**…**" Points and your total earned points are "**…**"

Thank you for completing the experiment.

Chapter 2:

Who is Irrational? The Endowment Effect and Concerns with Good-Dealness

“Not to be absolutely certain is, I think, one of the essential things in rationality”. - Bertrand Russell.

2.1 Introduction

In economics and psychology there are two central and widely used measures of valuation: willingness to pay (WTP) and willingness to accept (WTA). The WTP is the maximum amount of money an individual would pay to acquire a good and the WTA is the minimum amount of money an individual would accept to sell a good. These measures have been used in policy to provide valuations for non-market goods, consequences of public policies, and to derive compensation for losses. In the absence of income or transaction cost effects these two estimates are expected to be equal under standard economic theory, leading to indifference between the measures (Willig, 1976).

For example, based on the assumption that these two valuations would yield similar estimates, WTP elicitation has been solely and widely used to determine the gains or losses associated with an environmental improvement or an environmental catastrophe (Brown & Gregory, 1999). Nevertheless, much empirical evidence has shown that there is a disparity: WTA estimates exceed WTP estimates. Depending on the evaluation measure used, the benefit or the cost of a particular policy could therefore be underestimated or overestimated.

This disparity has most frequently between attributed to cognitive bias on the part of buyers and/or sellers, and it is typically assumed that owning an item changes the owner's perception of the value of that ownership. The aims of this chapter are to move beyond the traditional explanations that have been proposed to explain this disparity and to show that the endowment effect reflects strategic considerations rather than ownership-induced changes in underlying preferences. We develop an alternative model, based on a quantification of "good dealness", which assumes no cognitive bias or ownership-induced changes in underlying preferences. The model assumes that item ownership has no effect on beliefs about (a) the quality of the item (b) the appropriate market price for the item, (c) the market prices for other same-category items or (d) any ownership specific differences in participants' underlying valuations of the item. Instead, the model assumes that sellers rationally demand a market-appropriate price for the item given their beliefs about the item's relative quality and their beliefs about the distribution of market prices in the market. Buyers, in contrast, offer less than what they believe the appropriate market price to be because they typically do not want the item and will only offer a price that represents a good deal. We do this in a series of two experiments in which we find support for the model's key assumptions. The chapter is then structured as follows: we firstly discuss the literature on the endowment effect and proposed explanations, secondly, we discuss in detail our aims and motivations, thirdly we outline the details of each experiment performed, present the results, and then conclude with a discussion.

2.2 The Endowment effect as a Behavioural Bias

Thaler's study (1980) was the first to establish the term "endowment effect" to describe the disparity between WTA and WTP. This disparity, often numerically expressed in terms of the ratio WTA/WTP, has since become a poster child for the field of behavioural economics.

One of the first laboratory experiments investigating the endowment effect using a consumer good was Knetsch and Sinden's 1984 study. In their study, two experimental conditions were used in which all participants received a redeemable voucher. Half of the participants were asked if they were willing to accept $2 to give up the voucher and the other half was asked if they were willing to pay $2 to keep it. While among those who were asked to pay for the voucher the spread between those agreeing to pay and those not was equal, among those who were asked to accept the monetary amount only 24% accepted. It was argued by Knetsch and Sinden (1984) that valuing a potential loss leads participants to make different trade decisions about whether to trade then when valuing a potential gain.

In subsequent studies by (Knetsch, 1989) participants were found to be reluctant to trade their initial endowments of chocolate with mugs and mugs with chocolate but, when required to choose between a chocolate or a mug, the spread was fairly equal with 56% of participants choosing the mug. Similarly, in Kahneman, Knetsch and Thaler (1990) participants were reluctant to trade mugs that were initially endowed for pens and vice versa with Knetsch arguing that this result runs against the predictions of standard economics and the reversibility of indifference curves, -people should be choosing the good they prefer irrespectively of ownership.

Extensively studied over the last half-century (Horowitz & McConnell, 2002; Tunçel & Hammitt, 2014), the endowment effect has been observed with a plethora of goods, from mugs and chocolate bars to nuclear waste repositories, and its observation has been found to be robust to a large number of experimental techniques and designs (hypothetical or real payoffs; incentive compatibility; open-ended or closed questions; student or non-student participants), though the magnitude of the WTA/WTP ratio has been found to vary systematically with experimental parameters. For example, both market experience of valuing or transacting a good and experimental experience in a repeated design has been found to reduce the WTA-WTP disparity, but not eliminate it (Tunçel & Hammitt, 2014).

2.3 Traditional Explanation

One of the most influential accounts of the endowment effect is the work by Tversky and Kahneman (1991, 1992) who explained the endowment effect in terms of loss aversion, according to which individuals are hypothesized to weigh a loss higher than an equivalent gain when making a decision. This account was crystallized with the development of prospect theory, which incorporated an asymmetrical utility function relative to a reference point, thus formalizing the concept of loss aversion.

In prospect theory the reference point is a free parameter and typically being assigned to the status quo (although Kőszegi & Rabin, 2006 gave a formal expression for the reference point in terms of expectations and past experiences about transactions). Championed by Tversky and Kahnemann, the mantra that 'losses loom larger than gains' has since been cemented in the behavioural sciences lexicon with loss aversion being the traditional explanation for the endowment effect. This may seem unfortunate given that the valuation gap was empirically established much earlier (Birnbaum & Stegner, 1979), and also because loss aversion failed to offer much more than just a descriptive label for this particular behavioural phenomenon (Gal, 2006).

The loss aversion explanation for the effect is, of course, not the only one. Over the years, many additional cognitive, affective and motivational mechanisms have been put forward as possible underlying causes of the WTA-WTP disparity (Morewedge & Giblin, 2015). These accounts offer different origins related to where the bias comes from — is it sellers who behave irrationally, buyers, or both?

2.4 Who is Biased?

2.4.1 Sellers overvalue

In most accounts of the endowment effect, the origin of the WTA-WTP disparity can be laid at the feet of sellers, buyers or both. In their classic work, Kahneman, Knetsch and Thaler (1991) conducted a standard endowment effect experiment in which buyers and sellers were asked whether they would sell or buy a coffee mug at a range of prices from $0.25-$9.25. A third group, the choosers, could choose either a mug or money for each price. The only difference between sellers and choosers was the possession of the mug at the time of the decision; in every other respect their situations were identical. However, the choosers behaved as buyers, valuing the coffee mug at similar levels to buyers compared to the valuations of sellers. This result has been widely accepted as strong evidence that the endowment effect originates from a bias among sellers, who overstate the value of an object that they own due to loss aversion.

Beyond the mere loss aversion account, overestimation of an object's value is also consistent with the notion of psychological ownership (Morewedge & Giblin, 2015; Walasek, Rakow, & Matthews, 2015). Belk (1988) acknowledged that what we own constitutes a large part of our self-identity, and Pierce, Kostova and Dirks (2003) argued that it characterizes the human condition. Feelings of ownership have been found to have powerful psychological effects such as improving employee attitudes, organization and work behaviour as well as self-esteem and citizenship behaviour (Dyne & Pierce, 2003).

For example, Morewedge, Shu, Gilbert and Wilson (2009) extended the classic endowment effect paradigm by introducing groups of 'brokers' who acted on behalf of buyers or sellers. Buyers' brokers who owned a copy of the target item valued it more highly than did brokers who acted on behalf of sellers and did not own the target item (see also, Strahilevitz & Loewenstein, 1998). The results of their studies favoured the ownership account rather than the loss aversion explanation.

The concept of ownership was refined by Reb and Connolly (2007) who distinguished factual ownership and perceived ownership of an object (a chocolate bar or mug). They studied the endowment effect using a 2x2 treatment of possession/no possession and ownership/no ownership. They found that possession, rather than factual ownership, led to an endowment effect, indicating that the perception of ownership rather than the factual ownership of an item induced the endowment effect.

The accounts listed above suggest that owned objects are regarded as more valuable than non-owned possessions, and that elevated WTAs are a behavioural signature of this effect. In fact, it has been argued by Huck, Kirchsteiger and Oechssler (2005) that ownership improves one's position in bilateral trades leading to an evolutionary advantage.

Taken together, the endowment effect literature began, and in part continues, with the assumption that ownership status alters one's preferences.

2.4.2 Both sellers and buyers are biased

Other studies have suggested that both sellers and buyers are biased in producing the endowment effect. For example, Gal (2006) took a neutral position of whether buyers or sellers drive the endowment effect, arguing instead that a status-quo bias is ultimately responsible. In a separate and more recent vein, various researchers have shown that owners and non-owners differ in how they evaluate products. Carmon and Ariely (2000) argued that both buyers and sellers are biased. They theorised that individuals weigh more heavily things that they will forgo in the transaction, so that buyers emphasise expenditure whereas sellers emphasise giving up the item in question. To test this hypothesis they conducted several studies, finding that buying price was typically correlated with variables related to expenditure and selling price with variables related to ownership.

Alternative accounts have focused on the cognitive mechanisms of attention and perspective as indicative of differences in individual valuations. For example, Nayakankuppam and Mishra (2005), extending the work of Carmon and Ariely (2000), conducted experiments indicating that sellers focused more on positive features of an item and less on negative features than buyers did. The first of their three experiments asked buyers and sellers to list thoughts about the item (a pen) and indicate whether the thoughts were positive or negative. In the second experiment, they found that sellers were less likely to correctly remember negative facts about the item, but more likely to remember positive facts. Their third experiment showed that the endowment effect could be moderated by forcing sellers to pay attention to negative features and buyers to pay attention to positive features of the item.

An alternative theory of this type was discussed by Johnson, Haubl and Keinan (2007) who hypothesised, using a query theory of the endowment effect, that the difference between buyers and sellers is based on different valuation questions they ask themselves of the item, as well as the order in which these questions are asked. They conducted an experiment asking participants to list aspects of the item that led them to reporting their valuation (WTA/WTP). They found that choosers (buyers) produced more value decreasing aspects than sellers, whereas sellers produced more value increasing aspects than choosers. In a subsequent experiment, Johnson et al. (2007) were able to eliminate the endowment effect by forcing sellers to first consider value-decreasing aspects and buyers to first consider value-increasing aspects.

Further support for the bidirectional bias has been found in the context of information search. Pachur and Scheibehenne (2012) argued that buyers and sellers differ in their approach to information searching. They conducted a within-subjects experiment in which participants, buyers and sellers, were asked to determine the value of a lottery by experiential sampling. Their results indicated that sellers terminated their search after sampling high lottery outcomes, while buyers terminating their approach after sampling low lottery outcomes, thus producing the endowment effect.

Ashby, Dickert and Glöckner (2012) also used a lottery task, and contended that buyers and sellers differ in the allocation of attention to attributes of the item, arguing that sellers shift attention towards value increasing attributes, while buyers shift attention towards value decreasing attributes. Their theory predicted an increase in the WTA-WTP disparity with deliberation time. In their study participants were shown a series of lotteries, presented as a list of outcomes (all positive) and their associated probabilities. Participants were asked for their WTA/WTP for each lottery and were given a time limit of either 5, 10 or 15 seconds. They found that the endowment effect increased for studies with longer deliberation time. A further study included eye-tracking, allowing them to confirm that buyers spend more time and visual attention on value decreasing aspects of the item with sellers focusing on value increasing aspects. Ashby et al. (2012) also found that overall participants focused more on high probability outcomes.

Using a similar eye-tracking methodology Ashby, Walasek and Glöckner (2015) extended the investigation of Ashby et al. (2012) beyond monetary gambles to study common consumer goods (chosen from the Amazon.de marketplace) by measuring the relationship between attention to consumer ratings and valuations. Participants were simultaneously shown the average negative rating (between 0 and 2.5 stars) accompanied by the percentage of customers that gave a negative rating and the average positive rating (3.5 to 5 stars) along with the percentage of people who gave a positive rating. The measure of attention (Low-Gaze-Proportion, LGP) was equal to the duration of fixations on low rating components divided by the total gaze time on all negative and positive rating components. LGP predicted valuations, having a significant, negative relationship (even when information about customer ratings was altered to a frequential format to resemble the real Amazon marketplace). When participants were randomly assigned to take the perspective of either buyer or seller, buyers focused on lower ratings more than sellers, although both conditions fixated overall more on positive ratings than negative ratings which was partially attributed to the location of the positive information on the screen. Contrary to what is found with monetary gambles, the effect of perspective on attention in this riskless choice task mediated valuations significantly and negatively but to a smaller extent.

Finally, Plott and Zeiler (2005) have argued that the endowment effect is an artefact of the experimental design and that even the use of the words “sell” and “buy” can prime owners to sell high and non-owners to buy low. In a situation of strategic misrepresentation both sellers and buyers would be biased to not reveal their subjective valuations but increase and decrease them respectively. Removing these misconceptions in the lab by introducing training, paid practice, anonymity and incentive compatible elicitation method, Plott and Zeiler (2005) managed to eliminate the endowment effect (see Isoni, Loomes, & Sugden, 2011, for opposition to this viewpoint). Contrary to suggestions in Plott and Zeiler (2005), Fehr, Hakimov and Kübler (2015) and Bartling, Engl and Weber (2015) still found an endowment effect even for participants that understood the BDM mechanism and even when an easier price list incentivisation mechanism was used. Regardless of the incentivisation mechanism being applied, participants may maintain a habit from every-day life wherein they use the strategy heuristic (Korobkin, 2003) although evidence from experiments where participants are not willing to give up their initial endowment for other objects show that the endowment effect is still present.

2.4.3 Bias among buyers

More recently, some authors have suggested that it is buyers who often understate their WTP and that sellers' valuations are less biased. Buyers have been found to have a greater propensity for ambiguity aversion than sellers. In Trautmann, Vieider and Wakker (2008) one treatment group faced a choice between a risky and an ambiguous prospect and had to explain their decision to the experimenter, while the other treatment did not reveal any personal information. This led to fewer participants choosing the ambiguous prospect under the first treatment. In a secondary experiment, participants completed the FNE (fear of negative evaluation) scale (Leary, 1983) with the group that was more sensitive to negative evaluation exhibiting a bigger WTP difference (evaluation of the risky minus the evaluation of the ambiguous prospect) than did the less sensitive group. In the presence of FNE, WTP decreased (for ambiguous prospects) and led to fewer participants choosing the ambiguous prospect. Thus, the presence of ambiguity and imagined or experienced embarrassment can bias buyers' WTP.

Continuing this line of investigation, Trautmann, Vieider and Wakker (2011) also found higher WTP for risky prospects compared to the ambiguous ones, using a variety of experimental conditions. Nevertheless, in direct choice they observed preference reversals for a percentage of participants who preferred the ambiguous prospect to the risky prospect (although their WTP was higher for the risky prospect). They suggested that participants used the risky prospect as a stronger reference point and adjusted the WTP of the ambiguous prospect downwards. When participants were first endowed with these prospects and then asked for their WTA, only a minority demonstrated these preference reversals. Trautmann and Schmidt (2012) provided further evidence for the idea of Trautmann et al. (2011) that the risky prospect acts as a stronger reference point in WTP valuations than in WTA valuations and that the negative aspects of the ambiguous prospect weigh more heavily than the positives as compared to the risky prospect. In their study, WTP and WTA valuations were elicited for risky and ambiguous prospects under the comparative (two groups: buyers and sellers) and noncomparative (four groups: two buyers and two sellers, each group evaluating only one prospect) between-subjects design and within-subjects design (one group: both buyer and seller perspective). WTA valuations were higher than WTP valuations in all conditions, with ambiguity negatively influencing both WTA and WTP compared to the evaluations for risky prospects. In the comparative condition WTP was higher for risky prospects and lower for ambiguous prospects compared to the noncomparative condition. In the comparative condition, moving from the evaluation of risky prospects to the evaluation of ambiguous prospects decreased WTP more than the decrease of WTA, thereby increasing the WTA-WTP disparity under ambiguity.

A theory of loss-attention developed by Yechiam and Hochmann (2013) argued that a proposition with potential losses can have a positive impact on performance. In their study, participants were asked to choose between two lotteries of differing expected value. In an attention-depleted setting (where the lottery was presented as a secondary task) participants displayed higher performance, i.e. were more likely to choose the lottery with the greatest expected value, when the lotteries included losses. This loss-attention account can explain why the WTA for gambles is closer to their expected value than WTP, the potential losses of the owner increasing the attention assigned to the situation.

Yechiam, Abofol and Pachur (2017a) re-analysed two earlier studies (Ashby, Dickert and Glöckner, 2012; Pachur and Scheibehenne, 2012) and conducted several further experiments to examine whether sellers or buyers were more sensitive to changes in expected values of lotteries. They found that sellers showed a greater degree of sensitivity, or relative accuracy, to the expected value of the lottery except in situations with a long deliberation time. Relative accuracy was defined as the rank correlation between a series of elicited valuations and expected values. In addition, they found that sellers' valuations were typically closer to the expected value of the lotteries than buyers.

These results were strengthened in a meta-analysis of monetary lotteries. Yechiam, Ashby and Pachur (2017b) compared the absolute deviation of buying and selling prices for lotteries from their expected value. They also examined relative accuracy to account for risk aversion. Their analysis of 35 studies over the period 1967-2016 showed that sellers are more accurate, both in absolute and relative, terms than buyers. In addition, they found that sellers' valuations typically showed higher variance between participants than buyers' valuations did, though sellers exhibited a lower coefficient of variation (amount of variability relative to the mean). Importantly, Yechiam et al. (2017b) noted that while incentivisation reduced differences between buying and selling prices, it did not eliminate them.

These findings are in line with the predictions of the loss of attention account, which maintains that sellers engage in more cognitive effort than buyers (Yechiam & Hochman, 2013). Owners, who are motivated by the prospect of losing what they own, pay more attention to the task and provide valuations closer to the actual objective worth of an object. At least in the context of monetary gambles, where the objective value is known, the majority of results are consistent with this account.

2.5 Reference Price Theories

As evidenced by much of the previous discussion, there is no consensus as to the origins of the WTA-WTP disparity amongst current research, which is inconclusive in addressing the question of whether the WTA-WTP disparity is due to the behaviour of sellers, buyers or both. Furthermore, much of this research is limited to monetary gambles, and so does not generalise to other contexts, such as consumer goods. Indeed, Yechiam et al. (2017) explicitly state this in the interpretation of their work in other contexts.

There has been recent work (Isoni, 2011; Weaver & Frederick, 2012) arguing that the endowment effect originates in 'bad deal' aversion. A deal depends on reference prices in order to be framed as good, bad or neither. For example, Isoni (2011) suggests that the discrepancy between sellers' WTA and buyers' WTP reflects not differences in ownership underlying preferences for an object, but rather an aversion to a bad deal. Specifically, buyers are more averse to the possibility of overpaying for an object than they are to potentially miss out on the opportunity to buy the object if they fail to offer a high enough price. By taking shaping effects into account, whereby an individual's reference price depends upon recent observed prices, Isoni (2011) was able to move beyond the one-shot paradigm of most accounts of the endowment effect and incorporate the well-known reduction in the WTAP-WTP disparity in repeated trials into his theory.

Weaver and Frederick (2012) expounded a reference-price based theory of bad deal aversion. Specifically, they hypothesised that the WTA (WTP) is the maximum (minimum) of the item's reference price and the monetary value of their expected benefits from using the good (or consumption utility: Kőszegi & Rabin, 2006). This theory predicts that the endowment effect disappears as these two values converge, and predicts the existence of a characteristic U-shaped curve of the WTA-WTP disparity as a function of reference price near to this convergence (where the reference price is equal to the consumption utility). In their study, they tested this prediction in a series of experiments, in which participants in different treatments received different external reference prices for various items, including both regular goods as well as risky prospects (lottery goods) and were then asked to specify how much they would either pay or sell these items for. The results revealed that sellers' valuations were closer to the retail prices than those of buyers. For instance, in two conditions, a candy's price tag was presented to the participants as either \$4.00 or \$1.49. Mean buying price for the candy was largely immune to the change in the value on a price tag (being \$1.54 and \$1.20 respectively) but selling price was highly sensitive to it, being about 80% higher in the high compared to the low-price tag condition (\$2.88 compared to \$1.58). These studies obtained the predicted U-shape of the WTA WTP disparity as a function of reference price, though in the large-stakes lottery they examined the nadir of the U-shape did not correspond to the predicted equality of WTA-WTP. In their fourth study Weaver and Frederick (2012) attempted to vary consumption utility by using two chocolate bars of the same price but with different tastes. They found that WTP was heavily affected by the chocolate bar type, but WTA was not, consistent with the consumption utility but not the reference price being changed. In their final study they used the exchange of goods paradigm, endowing participants with one of two goods and asking if they would like to trade. In the high reference price condition, very few participants asked to trade, whereas in the low reference price condition more trades were observed. We note that in their review Ericson and Fuster (2014) argue that the theoretical approach based on bad-deal aversion and reference prices can be accommodated within a loss-aversion theory by setting the reference point as the reference price.

Based on these findings, Weaver and Frederick (2012) concluded that "Consumers evaluate potential trades with respect to salient reference prices, and selling prices (or trading demands) are elevated because the most common reference prices—market prices—typically exceed valuations." (p. 696).

Further evidence for bad deal aversion is provided by Brown (2005) who used the verbal protocol technique to understand individuals' motivations for their WTA and WTP valuations in a within-subjects experiment. His study found that the most commonly cited reason for giving both a high WTA and low WTP was a form of bad deal aversion, or attempt to obtain a good deal. Indeed, the verbal accounts mirrored almost exactly the theoretical motivations of Weaver and Frederick (2012), with participants' WTA reflecting their desire to not sell the item for less than its value in a sale situation (avoiding a bad deal). In contrast, WTP was typically based on a desire to obtain the item for a low price (seeking a good deal) or was based on what participants felt the item was worth to them (consumption utility, as per Weaver and Frederick, 2012). While as in Weaver and Frederick (2012), reference prices may be provided externally, for most common items people have some pre-existing knowledge or assumptions about the market.

2.6 Present Work

In this chapter we challenge the assumption that preferences are affected by ownership status. We propose instead that the behaviour of both buyers and sellers is driven by considerations of what constitutes a good deal for them (Isoni, 2011). More importantly, we attempt to move beyond the idea that sellers, buyers, or both sellers and buyers are "biased", because the relevant normative price (departure from which would indicate bias) is unclear. Indeed, in most previous experiments the relevant normative price (departure from which would indicate bias) is blurred and set by the experimenters either ex-post or ex-ante.

Instead, we elicit participants' individual perceptions of the quality of a product that they have the opportunity to buy or sell, and we also elicit their beliefs about the distribution of market prices for that product type. This enables us to determine what each participant believes the appropriate market price for a product to be. For example, a participant might believe that a coffee mug is high quality (e.g., at the 80th percentile of the quality distribution) and also believe (e.g.) that the 80th percentile of the distribution of coffee mug prices is \$6.50. We can therefore examine how each participant's WTA or WTP relates to their beliefs about the appropriate price for a product, as well as establishing whether ownership status influences perceptions of quality, perceptions of market price distribution, or both. Most importantly, we quantify the "goodness of a deal" in terms of the difference between the relative ranked position of product's price valuation within the relevant market price distribution and the relative rank position of its quality relative to other similar products. For example, a coffee mug at the 80th percentile of the quality distribution on offer at the 60th percentile price clearly represents a good deal. A 40th percentile (quality) coffee mug at the 70th percentile price does not. In the present paper, we set out to determine the relationship between what constitutes a good deal for a given product and a person's WTA or WTP for that product.

Our approach therefore contrasts with previous research which has focused on locating the source of bias in either sellers (Morewedge et al., 1997; Reb & Connolly, 2007; Strahilevitz & Loewenstein, 1998), buyers (Trautmann et al., 2008; Yang, Vosgerau, & Loewenstein, 2013; Yechiam et al., 2017b) or both (Ashby et al., 2015; Johnson et al., 2007; Nayakankuppam & Mishra, 2005; Pachur & Scheibehenne, 2012). Instead, we build on and extend recent suggestions that strategic considerations such as bad deal aversion may be important in explaining the endowment effect (Brown, 2005; Isoni, 2011). However, Isoni (2011) does not offer any direct evidence that perception of the target object's attributes (e.g. its quality and the price it would typically sold for in the marketplace) are uninfluenced by the ownership status. While both Isoni (2011) and Frederick and Weaver (2012) show that the WTA-WTP gap may be influenced by strategic considerations, neither study elicits participants' individual judgments of the appropriate market price for an object. This is important because it is possible that ownership status might influence participants' judgments of either the quality of an object or the appropriate price for that object. Such differences could occur if, for example, sellers focused on particularly positive attributes of the objects that they own or brought to mind higher prices when considering a reasonable selling price (Ashby et al., 2015; Ashby et al., 2012; Carmon & Ariely, 2000; Johnson et al., 2007; Nayakankuppam & Mishra, 2005; Pachur & Scheibehenne, 2012). Moreover, the findings reported by Frederick and Weaver (2012) could be explained by the fact that sellers exhibited more anchoring to (and insufficient adjustment from) the salient market value than buyers. This explanation is consistent with Simonson and Drolet (2004), who found that selling but not buying prices were influenced by anchors related to market prices and buying but not selling prices were influenced by arbitrary anchors such as the last two digits of one's social security number[12]. This account is also in agreement with a study by Sugden, Zheng, and Zizzo (2013), in which buyers were more sensitive to random anchors than sellers (though in other studies, the reverse pattern was found; e.g., Fudenberg, Levine, & Maniadis, 2012; Maniadis, Tufano, & List (2014).

Our objective in the present work is to re-evaluate the endowment effect in terms of buyers' and sellers' perceptions about quality and market prices. We offer new tests of the accounts that posit that the WTA-WTP valuation gap is not due to preference shifts, but rather reflects individuals' considerations of what constitutes a good deal. We achieve this by eliciting participants' beliefs about quality and market prices and examining how their valuations (WTAs and WTPs) relate to what they believe the product *should* cost in the market. In our approach, we elicit quality estimates not by asking participants to provide quality ratings on a 1-5 scale, but by asking them where they believe the product's quality ranks within the distribution of other members of that product category. This approach is motivated by considerable evidence that people base their judgements of quantities such as quality and price on the relative ranked position of the relevant quantity within a retrieved comparison set (e.g., Stewart, Chater, & Brown, 2006). In addition, rather than simply asking participants for their estimate of a typical market price for the object, we elicit from them their beliefs about the whole distribution of the market prices for that category product. This enables us to do two things. First, it allows us to predict what a participant should state as the appropriate market price for an object. For example, if they estimate a coffee mug to be at the 60^{th} percentile of the quality distribution, then (if they are behaving coherently) their judgment of the appropriate price for that object should be at the 60^{th} percentile of what they believe the market price to be. Second, this rank-based

[12] In their second study they asked participants to describe the motivations behind their valuations, with most of sellers basing their decision on the expected market price of the item.

methodology enables us to quantify the goodness or badness of a deal as described earlier.

Our objective is to determine what a person believes an object *should* cost in the market, based on that person's perception of the good's quality and its position in the broader market. We must therefore elicit a person's perceived market price distribution for a given product category. Such elicitation is challenging for two reasons. First, it can be difficult to avoid biased responding (e.g., anchoring effects) when eliciting distributions (Mazar, Koszegi, & Ariely, 2014). For example, Sharpe, Goldstein and Blythe (2000) elicit distributions by presenting participants with income ranges and requiring them to indicate how much mass of the relevant distribution falls within each interval. It is plausible that results would be strongly influenced by the location of the largest interval. Second, it is possible that the ownership status influences people's perceptions of either a) the market price distribution and/or b) a product's quality ranking. Our Experiment 1 is designed to determine that we can reliably elicit market price distributions with a minimal amount of error. It also allows us to establish whether owners and non-owners differ in their perception of the market's structure. To foreshadow: We find that our method works well and that buyers and sellers do not have different beliefs about market price distributions.

In Experiment 2, we complement the design of Experiment 1 by asking our participants about their perceived quality rank of the product that they were, or were not, endowed with. We are thus able to match people's perception of quality onto each individual's market price distribution and find what that person expects the object should cost. We also elicit the actual perceived market prices of the object, in order to determine whether it is true that these reference prices are higher than stated WTAs and WTPs (Weaver & Frederick, 2012). We also elicit confidence intervals for both price and quality in order to test the hypothesis that owners might be more confident about the quality and/or price of the relevant objects even if they did not differ in their single-point estimates. To foreshadow: our results show a remarkable degree of coherence in the behaviour of owners and non-owners. The two groups do not differ in their perception of market prices (as per Experiment 1) nor do they differ in their perception of the product's market price. Sellers and buyers of a product also did not differ in their perception of how highly the product ranks among other similar goods in terms of its quality. By matching people's personal beliefs about quality and market price, we found that this appropriate price for a product was very similar to sellers' valuations, but much higher than buyers' WTPs.

2.7 Experiment 1

2.7.1 Methods

2.7.1.1 Design

In a between-subject design, we compared valuations of owners (sellers) and non-owners (buyers) of a University branded water bottle. We also examined their beliefs about the market price distribution of the same object.

2.7.1.2 Participants

We recruited 79 participants using Warwick University's pool of volunteers (M_{age} = 20.70, 59% female). Each individual was promised a flat fee of £3.00 but were told that, depending on their choices, they could earn between £0.00 and £20.00 extra. Each session lasted approximately 30-40 minutes.

2.7.1.3 Procedure and Materials

Participants were tested in groups of maximum size 10. In any single session participants took on the role of buyer or seller. Sellers were given a brand new water bottle with the University of Warwick logo. These bottles were purchased from the University of Warwick bookstore (where their RRP was £6.99). Buyers were told that they would receive extra £4.00 (£7.00 in total with the flat fee of £3.00) for their participation. At this point, all participants completed an unrelated filler task and survey questions about this task that took on average 10 minutes to finish. Once done, sellers were reminded that they were given a water bottle and that it was theirs to keep. They were also told that they would have the opportunity to sell the water bottle if they so desired. Buyers were asked to examine the water bottle that the experimenter had just placed on their desk. They were told that they had the option of either buying the bottle and taking it home with them or keeping their money.

We used the BDM (Becker, Degroot, & Marschak, 1964) method to elicit people's valuations. Specifically, we informed our participants that a random price mechanism will be used to determine whether they would buy or sell the bottle. Examples of the BDM procedure were shown in both conditions; the examples were formulated abstractly in order to avoid any numerical anchors (see 2.14.1 Appendix I for the examples used in both conditions and instructions for Experiment 1). Participants were informed that at the end of the experiment the computer would generate a random offer/price for the water bottle. If the selling price was lower than that offer, then participants were required to sell the water bottle for the amount offered by the computer and were given that amount in cash. If the selling price was higher than the randomly generated offer price, then they were required to keep the bottle. In the case of buyers, if the buying offer was higher than that price, then participants were required to buy the water bottle for the amount chosen by the computer. If the buying offer was lower than the randomly generated price, participants kept their money and did not receive the bottle. On the subsequent screen, participants were asked the minimum price, in pounds, that they would be willing to sell their water bottle for. Buyers were asked the highest amount of money, in pounds, that they would be willing to pay for the water bottle.

After specifying their WTA or WTP for the water bottle, participants were asked about the prices of similar products in the market. Participants were shown an image of two rows of water bottles. They were asked to imagine that these bottles represent all unique bottles in the market and that they are ordered from the cheapest (leftmost bottle in the picture) to the most expensive (rightmost bottle in the picture). Participants were then asked to give their best estimate of the price, in pounds, corresponding to a specific position in the market. We used nine percentiles in total (10, 20, 30, 40, 50, 60, 70, 80, and 90). For each one, participants saw a red line corresponding to a particular position in the market (see Figure 2.1). They were then asked:

"The line indicates a price. [percentile]% of all water bottles cost less than the price indicated by the line, and [1-percentile]% cost more. What is the price indicated by the line (in British pounds)?."

Figure 2.1. Example question of the elicitation process that was shown to participants.

All nine questions were presented in a random order. The elicitation of the price distribution was incentivised so that the three individuals who gave the best (i.e. most accurate) estimates for prices of water bottles were awarded bonus payments of 15.00, 10.00 and 5.00 pounds for the first, second and third place respectively, after all sessions were concluded. The responses were compared to real price data for water bottles extracted from Amazon.co.uk. After the elicitation of the price distribution participants were reminded about the BDM procedure and that the computer would now randomly generate an offer/price and compare it with their selling/buying price/offer. The results of the BDM procedure were shown to participants, who were then asked to alert the experimenter. After all transactions were concluded, participants were thanked and debriefed.

2.8 Results

2.8.1 Exclusions

First, we removed one participant whose valuation (here WTA) was extremely high (> 4 * *SD* from the mean). We also identified and removed responses from one participant who did not provide consistent answers on the distribution elicitation task. More specifically, we calculated the Kendall Tau coefficient to determine whether people's responses were monotonically increasing with the percentiles of the distribution. We used a cut-off of 0.7 for our correlation coefficient, and we found that only one participant scored below this value (0.5). The final sample included 36 sellers and 41 buyers.

2.8.2 WTA/WTP Ratio

We found clear evidence of an endowment effect: Our sellers demanded on average £5.29 (*SD* = 3.97) for the bottle whereas buyers were willing to pay only £2.01 (*SD* = 1.92) to obtain it. The average WTA/WTP ratio of 2.64 is similar to that obtained in the endowment effect literature (Horowitz & McConnell, 2002; Tunçel & Hammitt, 2014). The values of WTA and WTP were determined as significantly different by both a two-tailed t-test, t (75) = -4.72, p < .001, *ci* = [-4.67, -1.90]), as well as a Mann-Whitney U test confirming that the endowment effect was found, Z = -4.52, p < .001, r = 0.51.

2.8.3 Distribution Elicitation Task

Figure 2.2 below summarises the mean and median responses on the distribution elicitation task. It is evident that owners and non-owners tended to be in agreement with regards to the market price distribution of the water bottles. We tested this with a MANOVA, in which we found no significant effect of ownership status on people's estimates of market prices (λ = 0.93, p = 0.82). A MANOVA with log-transformed values produced similar results (λ = 0.91, p = 0.71). We also conducted t-tests between each percentile's valuations for the two groups, with no significant differences being found.

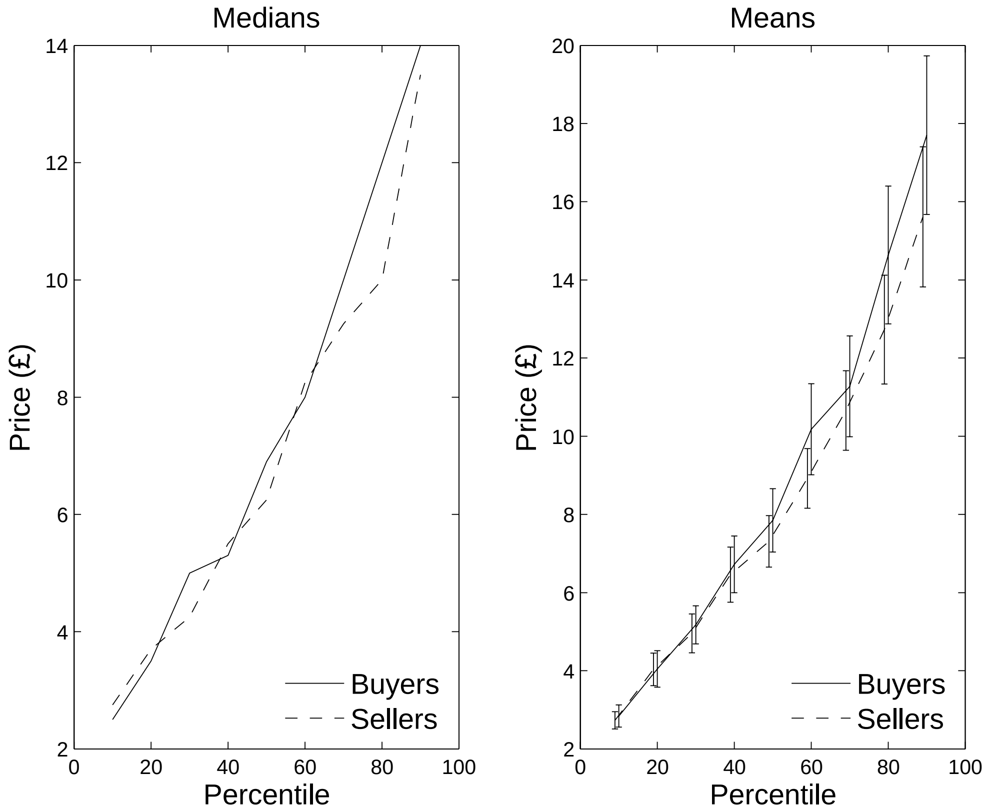


Figure 2.2. Median (left) and mean (right) percentile estimates of the market price in Experiment 1. Error bars in the right panel represent +- 1 standard errors of the mean.

2.8.4 Distribution fitting

What percentiles of the market price distribution do participants' WTAs and WTPs correspond to? In order to compute the rank position of each participants' WTA(P) in their elicited distribution of market prices, we fitted a lognormal distribution to each individual's responses (see 2.14.2 Appendix II for results using a normal distribution). In the fitting process, we used a least-squares parametric fit to lognormal inverse CDF functions. For each fitted distribution, we obtained the rank of each participant's WTA(P). All R^2 values obtained in the fitting are plotted in Figure S2 (see 2.14.2 Appendix II), and show that the quality of the fit was very good (M = 0.93, SD = 0.05).

We found that the average rank position of seller's WTAs was considerably higher (M = 0.34, SD = 0.26) than the rank of buyers' WTPs (M = 0.08, SD = 0.15), which is unsurprising given that we observed a large endowment effect in the absence of any differences in elicited market price distributions. Indeed, we found a significant difference between owners' and non-owners' ranked valuations, Z = -4.88, p < .001, r = 0.56. All ranks are plotted in Figure 2.3**.** First, it is clear that many owners and non-owners provided a valuation that ranks very low in their perceived market price distribution of all water bottles. This is particularly apparent among buyers the majority of whose valuations rank extremely low in the respective market price distributions. Among sellers, the price demanded for a water bottle corresponded to a wider range of positions.

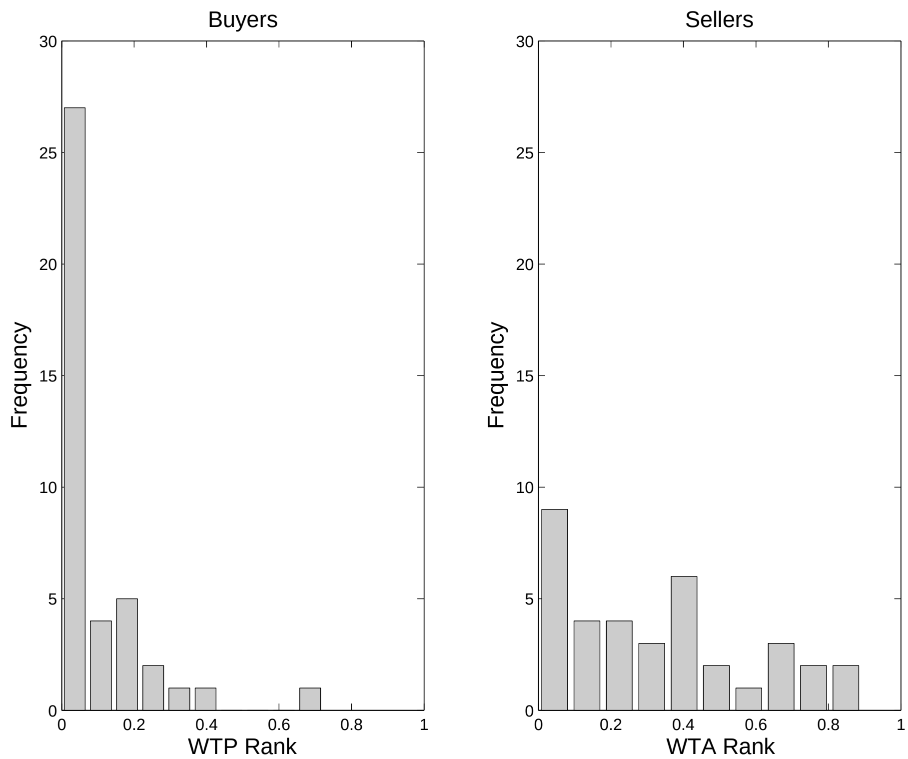


Figure 2.3. Histograms of WTA(P) ranks within individually fitted market price distributions.

2.9 Discussion

In our first experiment, we validated our methodology in two respects. First, we replicated the classic endowment effect with an incentive compatible valuation protocol. Second, we showed that our distribution elicitation method works well, with only one participant out of 78 failing to provide us with monotonically increasing market prices. Our results also showed us that buyers and sellers do not differ in their perception of the market prices. That is, for none of the percentiles that we used have we found that owners overstate the prices of similar products. The results therefore provide the first evidence that ownership status does not lead to a distorted perception of market prices, and hence that the endowment effect cannot reflect any such bias. As a result, valuation of buyers and sellers corresponded (here ranked) differently in their elicited beliefs about the spread of the market prices for water bottles. Whereas sellers' valuations reflect a wide range of possible market prices (see Figure 2.3), buyers treat the product as worth very little in terms of its market worth.

We extend this methodology in Experiment 2 to evaluate people's beliefs about the market price of the good, and their perception of its quality. This allows us to quantify the degree to which buyers and sellers differ in their concern with making a good deal.

2.10 Experiment 2

Experiment 2 had three overarching aims. The first aim was to further examine whether ownership status influences people's perception of objects' attributes. Specifically, we elicited participants' judgments of both quality and market price of the objects. We were also interested in whether participants' certainty about market price or quality might be influenced by ownership status. For example, owners and non-owners might give the same single point estimates for an object's quality, yet owners might be more confident about their estimate, in which case their 90% confidence interval for their estimate would be smaller. Our second aim was to determine whether participants' judgments of the market price of an object can be predicted from that person's judgment of the ranked quality of the item in combination with their beliefs about the distribution of market prices. Our third aim was to understand how people's WTAs and WTPs relate to their estimates of the quality and market price of the relevant object. In particular, we wanted to be able to quantify the amount of "good-dealness" required by buyers to purchase an object that they likely have little desire for, and the amount of "bad-dealness" acceptable by sellers when giving up the item.

2.10.1 Methods

2.10.1.1 Design

The design of Experiment 2 was identical to Experiment 1 but with the addition of quality rank and market price questions for the specific water bottle participants were given, and incorporated counterbalancing (between these two sections) to screen for order effects (see 2.14.3 Appendix III for the ANOVA results which found no strong evidence of interaction effects between the two orderings). These additional questions followed the elicitation of the market price distributions of water bottles.

2.10.1.2 Participants

We recruited 92 participants (M_{age} = 21.90, 63% female) from the University of Warwick, who were tested in groups of no more than 10. Each individual was promised a flat fee of £3.00 but were told that, depending on their choices, they could earn between £0.00 and £20.00 extra.

2.10.1.3 Procedure and Materials

The procedure was the same as in Experiment 1, but extended to accommodate new measures. Following the market price elicitation task, participants were asked to specify their best estimate of the actual market price for the target water bottle. To measure the uncertainty in their price estimates, participants could also specify an upper price (that they were 90% certain the true price was below), and a lower price (that they were 90% certain the true price was above). The exact question wording can be found in 2.14.1 Appendix I). Next, we elicited estimates of the quality of the water bottle. Participants were shown a new picture of all water bottles in the market, but now ordered by their quality, starting from the most low-end (left) to the most high-end (right) (see Figure 2.4). Participants were told that at the high-end water bottles had the most features and best materials and at the low-end they had the fewest features and poorest materials. They were then asked to indicate, by clicking on the appropriate region of the scale, where they believed the water bottle that they were given (or were offered) ranked in terms of quality.

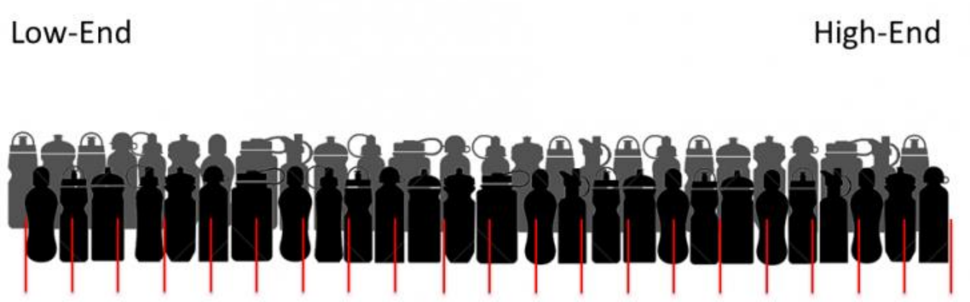


Figure 2.4. Quality rank question as shown to participants.

The rank participants gave was represented by a green rectangle on the graphic. After providing their single point estimates participants were asked to give low and high estimates of the water bottle's quality such that they were 90% sure the water bottle fell above the low estimate and fell below the high estimate. These estimates were also made using the graphical interface shown in Figure 2.4.

2.11 Results

2.11.1 Exclusions

We excluded two participants based on the same criteria as in Experiment 1. None of the participants provided us with extreme WTA/WTP values, but responses of two individuals were removed due to poor consistency in the distribution elicitation task (Kendall Tau < 0.7). We additionally excluded three participants who provided extreme values in the market price estimation question (> 4**SD*). In total the sample consisted of 44 sellers and 43 buyers.

2.11.2 WTA/WTP Ratio

As in Experiment 1, we found a clear endowment effect, with WTAs of sellers (*M* = 5.07, *SD* = 3.41) exceeding WTPs of buyers (*M* = 2.87, *SD* = 2.30). The WTA/WTP ratio of 1.77 was smaller than the 2.64 found in Experiment 1, but the difference between the groups was significant: two-tailed t-test, t (85) = -3.53, p < .001, ci = [-3.45, -0.96], as well as a Mann-Whitney test, Z = -3.20, p = 0.001, r = 0.34.

2.11.3 Market Price Analysis

Despite the differences in WTA and WTPs buyers and sellers did not produce different estimates of the market price for the water bottle or of the 90% confidence intervals for that price (see Table 2.1 below).

Table 2.1. Descriptive statistics (means) for the market price estimates of the water bottle.

Condition	N	Market Price	Standard Deviation	Lower Market Price	Upper Market Price
Buyer	43	5.57	2.68	3.40	8.58
Seller	44	6.07	3.07	3.73	10.68

There were no significant differences for either the market price value (t (85) = -0.80, p = 0.43, ci = [-1.72, 0.73] and Z = -0.80, p = 0.63, r = 0.09), or for the difference between the upper and lower end of the possible market price value (t (85) = -1.95, p = 0.05, ci = [-3.58, 0.03] and Z = -1.57, p = 0.12, r = 0.17).

2.11.4 Distribution Elicitation Task

Next, we examined responses from the market price distribution elicitation task. Mean and median responses on the market price distribution task are plotted in Figure 2.5. In line with the results of Experiment 1 there appear to be no differences between the estimates produced by owners and non-owners of the water bottle. Indeed, we conducted a MANOVA, in which the effect of ownership status was not significant, λ = 0.89, p = 0.390. A MANOVA with log-transformed values also produced a non-significant result (λ = 0.8947, p = 0.442). T-tests between each percentile's values among buyers and sellers also revealed no differences.

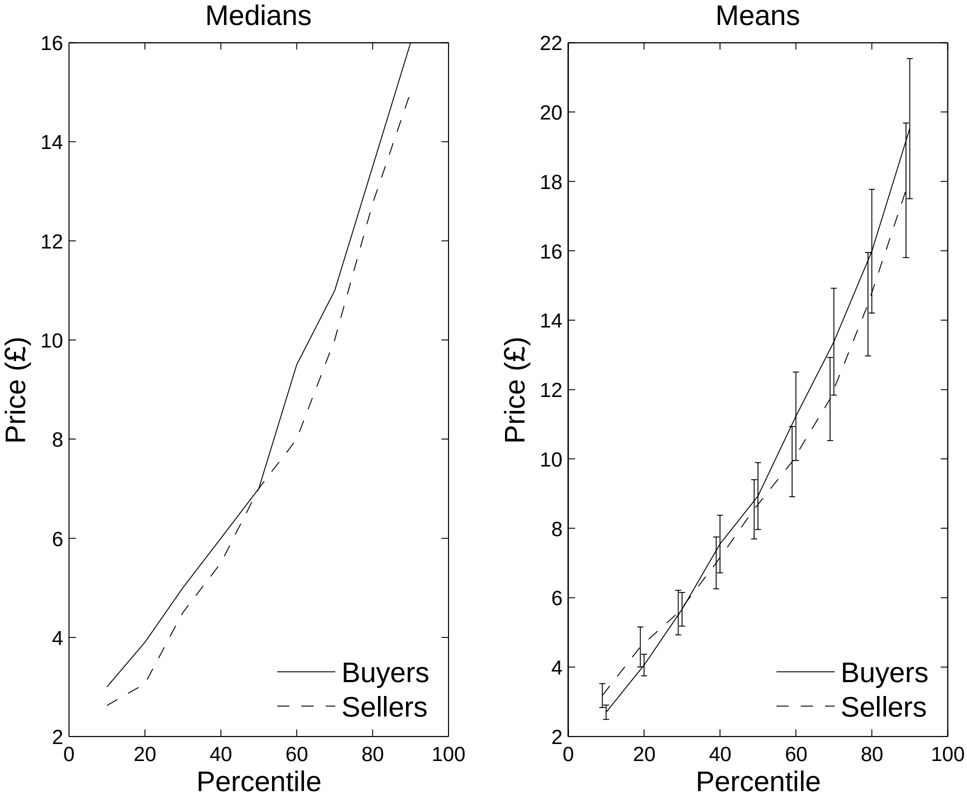


Figure 2.5. Median (left) and mean (right) percentile estimates of the market price in Experiment 2. Error bars in the right panel represent +- 1 standard errors of the mean.

2.11.5 Quality Rank Analysis

We now turn to the analysis of judgments of the target bottle's quality rank. Table 2.2 summarises the data for the average judged quality rank and for the upper and lower ends of the participant's 90% confidence cut-offs.

Table 2.2. Descriptive statistics (means) for the quality rank estimates of the water bottle.

Condition	N	Quality Rank	Standard Deviation	Lower Quality Rank	Upper Quality Rank
Buyer	43	0.2820	0.1510	0.1703	0.4459
Seller	44	0.3364	0.1794	0.1841	0.5614

There was no significant difference between the single point estimates of quality rank produced by buyers and sellers (*Z* = 1.32, *p* = 0.19, *r* = 0.14). We also examined whether buyers and sellers differed in the uncertainty of their quality rank estimates, which we define as the difference between the upper and lower quality ranks. We found a significant difference (*Z* = -2.39, *p* = 0.02, *r* = 0.26) such that sellers expressed a lower level of certainty about the water bottle's quality rank. More specifically, owners and non-owners produced similar estimates of the lower end of the quality rank confidence interval, but owners gave a higher estimate of the upper end of the range. Although the effect is small in magnitude, this result provides evidence against the hypothesis that ownership status increases owners' confidence in their estimates of the quality of the relevant product. In summary, there was very little difference between buyers' and sellers' estimates of either the quality of the product or its market price, indicating that the substantial endowment effect that we observed cannot reflect such differences.

2.11.6 Distribution Fitting

As in Experiment 2, we fitted a lognormal distribution to each participant's responses on the market price elicitation task (Please see 2.14.2 Appendix II for results using a normal distribution, Figure S3 in 2.14.2 Appendix II shows an overall high quality of model fits in terms of R^2 (*M* = 0.95, *SD* = 0.05)). We then determined the rank position of an individual's WTA or WTP in the fitted distributions. As in Experiment 1, we found that the ranked position of sellers' WTAs (*M* = 0.31, *SD* = 0.28) was different from the ranked position of buyers' WTPs (*M* = 0.14, *SD* = 0.20), *U* = -3.27, *p* = 0.001, *r* = 0.35. The distributions of ranks are plotted in Figure 2.6. Once again, we can see that both groups tended to provide valuations that ranked very low in the market price distribution. However, many more sellers than buyers provided valuations corresponding to market positions higher than the 10^{th} percentile.

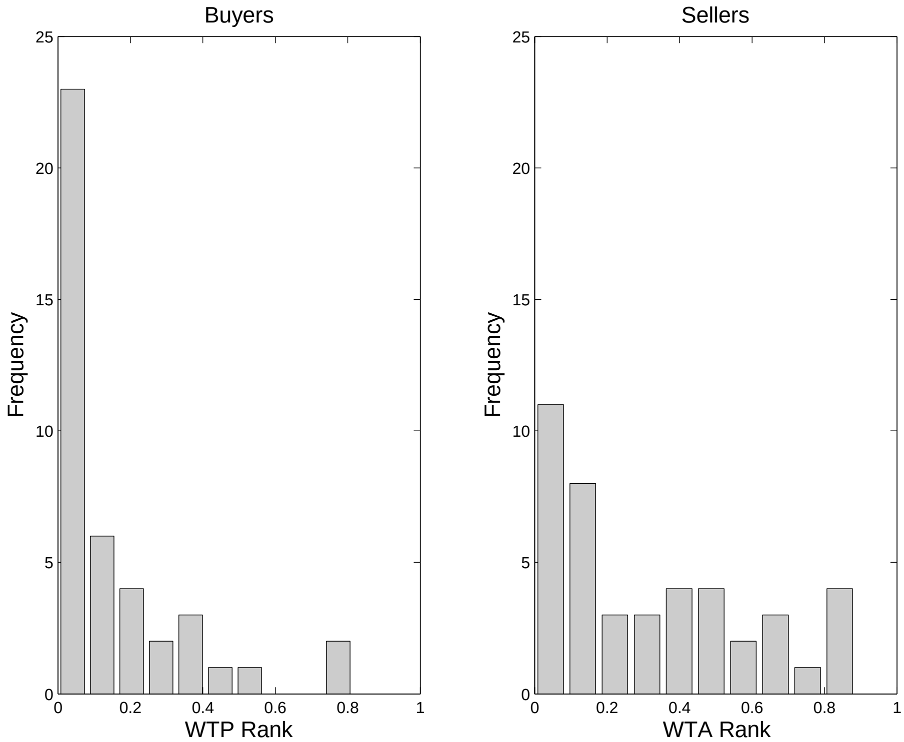


Figure 2.6. Histograms of WTA(P) ranks in Experiment 2.

2.11.7 Quality Matched Prices

Our results so far showed that owners and non-owners do not differ in their perception of the global distribution of market prices for a given class of consumer goods. In addition, ownership status does not seem to dictate people's perception of quality or perception of the object's market value. None of these quantities can, at least in isolation, explain why we observe the WTA-WTP gap. If we assume that participants believe the market price of the item is a monotonically increasing function of its quality (so that if a quality of an item increases each market price increases), then the rank of the item's quality is the same as the rank of its market price. We can now use each individual's estimates of the quality of the water bottle to quantify the good-dealness of the price they are willing to pay or accept. This is done in the following way. For each participant, we take their estimate of the ranked quality of the water bottle and find the market price that occupies the corresponding rank position in the distribution that represents that participant's market beliefs. For example, if a person stated that the water bottle was at the 30th percentile for quality, we calculate the 30th percentile price in the distribution of market prices that we elicited from that participant.

We refer to this estimate as the quality matched price (QMP). In the same fashion, we obtained the high quality matched price (HQMP), and lower quality matched price (LQMP), which correspond to the lower and upper bounds of the confidence intervals provided by the participant. More formally, if $F^{-1}(e; \mu, \sigma)$ is the fitted lognormal inverse CDF for a participant's elicited market price distribution, where e is the percentile and (μ, σ) are the parameters for the lognormal distribution, then the quality matched price, QMP, is related to the quality rank QR, by: $QMP = F^{-1}(QR; \mu, \sigma)$. Summary statistics for the three measures of quality matched prices are shown in Table 2.3. The quality price is a measure of participant's perceived market price of the item, given only the rank estimate for the quality of the item. It is what the participant perceives to be an *appropriate* market price for the item of that specific quality.

Table 2.3. Descriptive statistics (means) for appropriate prices.

Condition	N	Average Quality Price	Standard Deviation	Average Low Quality Price	Average High Quality Price
Buyer	43	5.30	3.09	3.48	8.46
Seller	44	5.71	2.86	3.99	10.86

Since we have not found any differences in quality ranks or market price distributions, it is unsurprising that we do not find any differences for the appropriate prices (U = -0.66, p = 0.51, r = 0.07) or the HQMP-LQMP range (U = -0.46, p = 0.64, r = 0.05) between sellers and buyers.

2.11.8 A Tale of Three Prices

For each participant, our analysis gave us three prices for the water bottle: the offer price WTA(P), the appropriate price and the market price. It is evident that the estimates of the appropriate prices are very similar to participants' estimates of the market prices of the product (see Table 2.1). Specifically, considering first buyers, the range of market prices estimated directly by the participants was from 3.40 to 8.58.

The corresponding estimates obtained from the relative rank matching procedure just described were 3.48 and 8.46. A similar pattern is seen for sellers: direct estimates were 3.73 and 10.68 for the lower and upper ends of the confidence intervals respectively; the corresponding estimates obtained by relative rank matching were 3.99 and 10.86. There was therefore a high degree of coherence in participants' estimates; their direct estimates of the market price for the water bottle were highly predictable from their estimates of the water bottle's quality combined with their beliefs about the market price distributions of water bottles' prices.

We summarise the results in Figure 2.7. This figure shows, for both buyers and sellers, three different prices. One is the WTA or WTP; the second is the participants' estimate of the market price for the water bottle, and the third is the estimate of the appropriate price as inferred from the quality matching procedure.

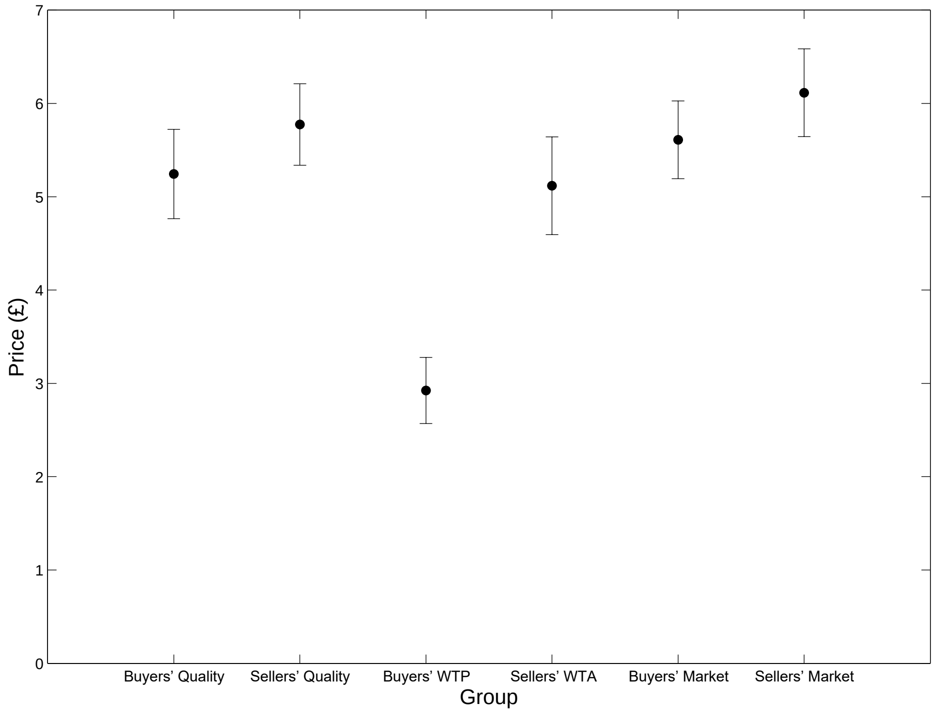


Figure 2.7. Valuations of buyers and sellers together with elicited market prices and estimated appropriate prices for the water bottle. Error bars represent +/- 1 standard errors of the mean.

All price estimates made by both buyers and sellers were within statistical error of each other except for a single price: the buyers' WTP. This suggests that it is the buyers who, in this experiment, are driving the endowment effect, as we discuss below. Note that all the prices are between £5 and £6, with exception of WTP, which is at £2.87.

While comparisons within price category (WTA(P)/Quality/Market) have already been performed in this chapter, we wish also test for differences between these categories. We therefore ran a two-way mixed ANOVA with buyer/seller as a between factor and price type (WTA/WTP, appropriate price and market price) as a within (repeated measures) factor, with the output shown in Table 2.4 below.

Table 2.4. Two-way repeated measures ANOVA results.

Variable	df	*F*	*p*	Partial η^2
Ownership status	1	3.87	0.05	0.044
Error (Owner status)	85			
Price Type	2	22.09	<.001	0.206
Interaction term	2	5.88	0.003	0.065
Error (Price Type)	170			

The results of the ANOVA show a strong effect of Price Type, as well as a strong interaction effect and a significant effect of the Ownership status (the means of these groups are plotted in Figure 2.7 above). To check if buyers' WTPs are responsible for the significant ANOVA results we conducted pairwise t-tests, using Bonferroni adjustments (there are 15 pairwise t-tests, so we find 0.05/15= 0.003 as a Bonferroni adjusted significance level). The t-tests were paired where necessary and unpaired otherwise. The results, shown in Table S10 in 2.14.5 Appendix V clearly indicate that buyer's WTP is significantly lower than all the other price estimates we obtained, with differences between all other estimates not being significant.

2.11.9 Appropriate Price Distribution

We can extend the analysis of where WTA(P) ranks in the market price distribution (as in Figure 2.6) to see where participant's WTA(P) ranks in an effective distribution of appropriate prices for the item. Using the three quality prices, LQMP, QMP and HQMP that we obtained we can form an *appropriate price distribution* (APD) for each participant. This distribution characterises the range of prices that each participant considers to be reasonable for the item, given its quality.

The APD is defined in terms of the three data points corresponding to the LQMP, the QMP and the HQMP. The LQMP and HQMP must correspond to rank positions 0.1 and 0.9 in the APD respectively. This is because low (high) quality rank estimate was elicited by asking for a quality estimate the participant was 90% sure the true value lay above (below). It follows from this that the LQMP and HQMP must correspond to rank positions 0.9 and 0.1 in the APD, respectively. The quality estimate itself was elicited as participants' estimate of the quality of the water bottle – it follows that the QMP must correspond to a statistical average of the APD. We assume that the elicited quality rank corresponds to a median value, so that the median (rank of 0.5) of the appropriate price distribution is QMP.

We then construct the CDF of the APD in a piecewise linear fashion, matching the (price, rank) values of (LQMP,0.1), (QMP,0.5) and (HQMP,0.9), so that the APD CDF consists of straight lines between these data points as well as the point (0,0). We also need to include a maximum price (MP) for the distribution, so that we can assign ranks in the APD to WTA(P) values above HQMP. For the maximum price, we take the price corresponding to the 99th percentile from the individual lognormal fits, and have it correspond to a rank of 0.99 in the APD. If the WTA(P) value is above MP, then we set its rank to 1.

The resulting piecewise linear CDF for the APD is illustrated schematically in Figure 2.8. For example, if in Figure 2.8 we insert participant's WTA(P) into the x-axis such that it is between QMP and HQMP, then their rank in the APD will be between 0.5 and 0.9. Similarly, if their WTA(P) is between 0 and LQMP then their rank in the APD will be between 0 and 0.1. For analytic numerical computations of the ranks of WTA(P) in the APD please refer to 2.14.4 Appendix IV.

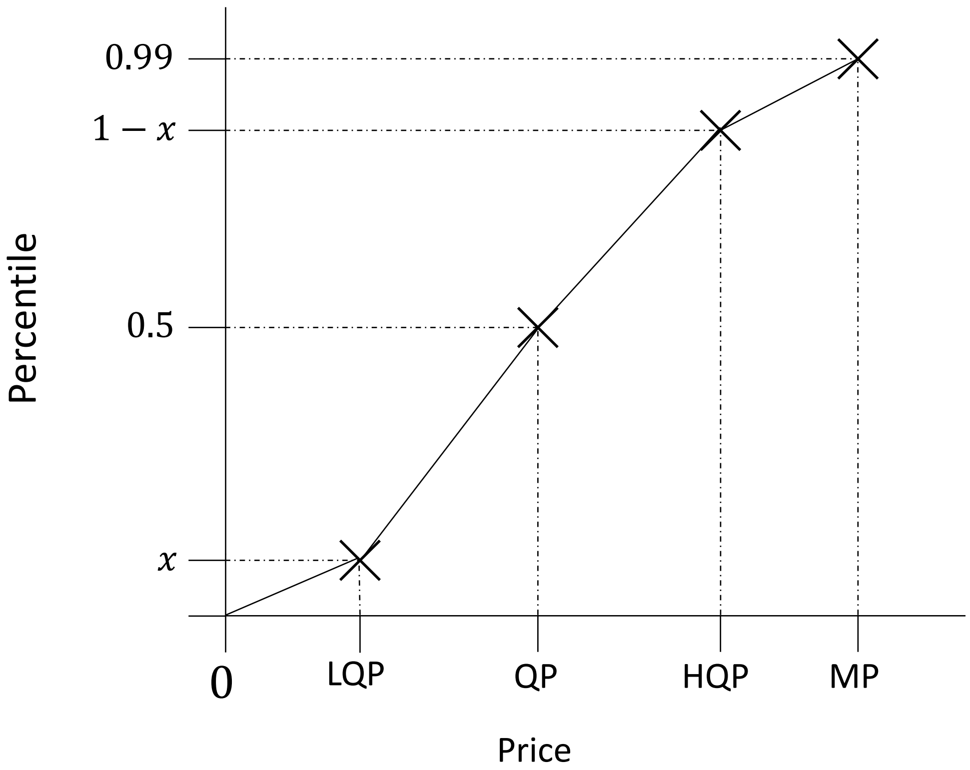


Figure 2.8. Schematic representation of the construction of the APD.

Once we have the APD, we calculate an appropriate rank for the WTA(P) of each participant, by finding the rank position of WTA(P) in the APD. Conceptually, this corresponds to the rank position of WTA(P) in each individual's distribution of appropriate prices for the particular water bottle in the study. We find the 'appropriate price rank' for buyers and sellers by inserting the WTA/WTP into the appropriate price distribution (APD) that we have just constructed and the result is shown in Figure 2.9.

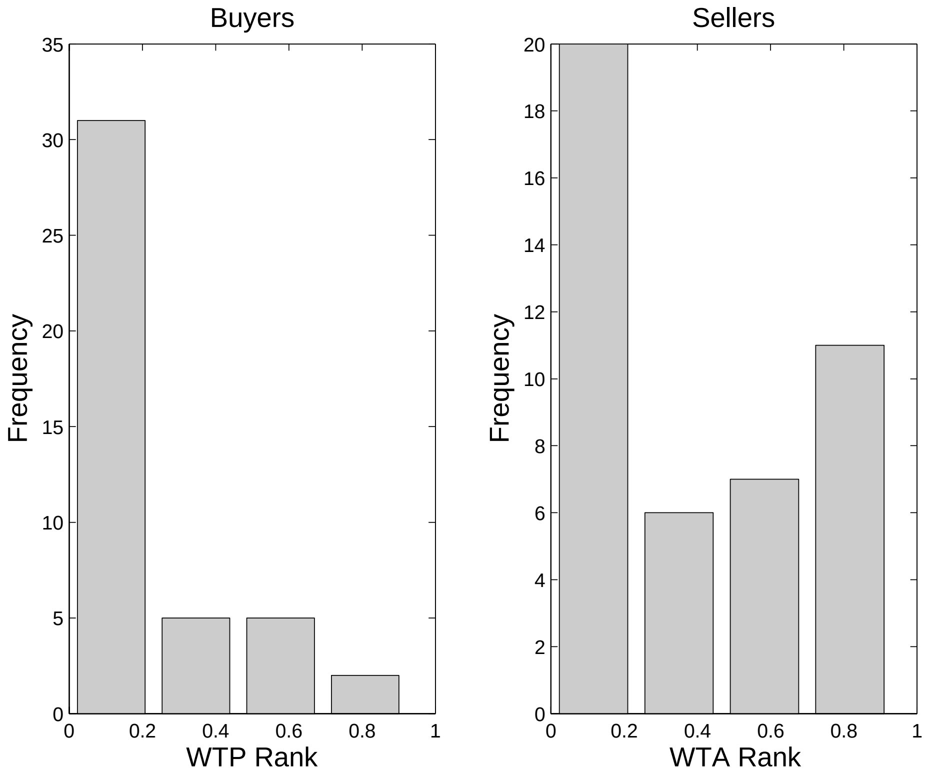


Figure 2.9. Histograms showing the distributions of appropriate ranks for buyers and sellers.

We can see that a lot of ranks are close to zero, though there are many lower ranks for buyers than sellers, as one would expect. Aggregate statistics are detailed in Table 2.5.

Table 2.5. Aggregate statistics for appropriate ranks.

	Average Rank	Std. Dev.	Median
Buyers	0.19	0.25	0.07
Sellers	0.40	0.35	0.33

The appropriate ranks for buyers and sellers were significantly different: $U = -3.02$, $p = 0.003$, $r = 0.32$. From this analysis, we observe that participants' WTA and WTP rank significantly different in their appropriate price distributions, with buyers having a lower rank position on average than sellers. This finding is similar to the results reported previously in Figure 2.6, namely that participants' WTA and WTP rank significantly differently in their elicited market price distributions (with buyers having a lower rank position on average than sellers).

An alternative analysis examines whether WTA(P), which we denote below by x, was lower than LQMP, between LQMP and QMP, between QMP and HQMP or above HQMP. The numbers of participants who fall into these four groups are reported in Table 2.6.

We observe many more buyers than sellers having a WTP lower or equal than LQMP and more sellers having a WTA between QMP and HQMP and above HQMP. This means that buyers often provided WTP that was within a range suitable for an item of lower quality. Sellers' WTA were more equally distributed so that some chose a price corresponding to a low-quality product, or even of a high-quality product.

The spread in appropriate ranks for sellers in particular suggests that the simple explanations in terms of aggregate quantities are perhaps incomplete. Nevertheless, the strength of the aggregate analyses demonstrates their validity. There are 18 sellers that give a WTA lower than LQMP, but the average LQMP for sellers is 3.99 which is still greater than buyers' average WTP (2.87), so these individuals might still contribute to the endowment effect.

Table 2.6. Number of participants with WTA(P) in each section of the APD.

	x<=LQP	LQP<x<=QP	QP<x<=HQP	x>HQP
Buyers	29	7	5	2
Sellers	18	8	8	10

We would expect an average appropriate rank equal to 0.5 if WTA/WTP matched perfectly the appropriate price. Our analysis of the appropriate price distribution revealed two features of people's valuation. Many owners and non-owners valued the consumer good as if it was of lower quality. However, considerably more sellers than buyers specified WTAs that correspond to the price of an object that is of a median or even higher than median quality.

2.12 General Discussion

In the present work, we explored the hypothesis that the behaviour of both buyers and sellers can be understood as rational given good-dealness considerations rather than reflecting any bias or irrationality on the part of either owners or non-owners of the object. Building on the argument that the endowment effect may reflect strategic considerations (Brown, 2005; Isoni, 2011; Weaver & Frederick, 2012), we investigated whether the endowment effect could be understood in terms of rational consumers' concerns with goodness and badness of a deal in the light of their perceptions of (a) quality of the item and (b) the distribution of market prices for such items.

In Experiment 1, we set out to determine whether we can reliably elicit people's beliefs about the market price distribution of a particular class of consumer goods, and whether these beliefs differ as a function of ownership status. Our experiment confirmed that it is possible to elicit coherent estimates of price distributions but we also found that these beliefs did not differ between owners and non-owners. In Experiment 2, we replicated these findings and also found that owners and non-owners did not differ in their estimates of the product's quality (in terms of its rank among other similar products). Moreover, owners and non-owners produced similar estimates of the product's actual market price.

For each participant, we estimated the price that individuals believed to be appropriate for the relevant object given (a) that participant's estimate of the ranked quality of the object and (b) that participant's estimate of the distribution of market prices of such objects. Our results can be understood in terms of sellers' rational desire to achieve something close to the appropriate market price and buyers' rational desire to purchase an item which they probably do not want only if they can do so at the price that represents a good deal.

For example, in Experiment 2, sellers estimated the water bottle to be at the 34th percentile of the quality distribution, and required a price at the 31st percentile of the price distribution to sell it. Buyers estimated the water bottle to lie at the 28th percentile of the quality distribution, yet were prepared to pay only the 14th percentile price. Thus, buyers were only prepared to offer a price that represented a good deal for them, where good-dealness is quantified in terms of the difference between the quality rank and the price rank.

We therefore suggest that the endowment effect can be naturally explained in terms of the different strategic considerations that apply to buyers and sellers. In a typical experiment, most participants will most likely not particularly want the relevant object (cf. Weaver & Frederick, 2012). Consider first the position of sellers. It is natural to assume that sellers, even if they anticipate little consumption utility from an object themselves, will be cognizant of the fact that somebody is likely to be willing to pay something close to the typical market price of the object. Imagine that you live in Australia, have a morbid fear of travel, and are left an apartment in New York by an elderly aunt. The consumption utility of the New York apartment for you is very low, but it is obvious that you would nonetheless put the apartment on the market for something near the appropriate market price, which (assuming efficient markets), is in itself an indication of what someone who wants that apartment would be willing to pay for that apartment. Clearly, a potential buyer of the apartment is in a different position. Assume now that a buyer, like a seller, anticipates relatively little consumption utility if they owned the apartment. The buyer will only be prepared to pay anything at all for the apartment to the extent that it constitutes a good deal that could potentially lead to a profit-making sale, and in the case of objects typically included in laboratory studies of the endowment effect such considerations may in any case largely be outweighed by transaction costs.

Our account differs from previous explanations of the endowment effect in several key respects. Most importantly, unlike traditional accounts based on concepts such as loss aversion, our account does not assume ownership-induced changes in people's underlying valuations of the object if such valuation is defined in terms of consumption utility (rather than, for example, the profit that could possibly be made by selling it). In this respect, our account is similar to that of Isoni (2011). However, unlike Isoni, we do not need to assume "bad deal aversion" in that we do not assume any asymmetry in hedonic impact of under- and over-paying.

Our results are comparable to the findings reported in the endowment effect studies of risky and ambiguous gambles. Sellers, not buyers, tend to set the minimum selling price to be close to the actual objective worth of a risky asset (Yechiam et al., 2017a; Yechiam et al., 2017b). Although in the present study we can only approximate what the worth of a consumer good is for each person, all of our findings are consistent with the fact that sellers' behaviour reflects their subjective view of what the item *should* be worth. Our results therefore extend previous efforts beyond the context of gambles.

Our results have wider implications concerning the use of incentive compatible procedure like the BDM (Becker, Degroot, & Marschak, 1964) to elicit true valuations. If the amount that people are willing to sell or buy an item for reflects strategic considerations relating to appropriate prices for an item of that quality, rather than an individual's desire to possess the object, the valuations obtained using BDM-like procedures cannot be interpreted as measures of preferences.

We note that the extent to which the strategic considerations based on beliefs about the market come into play will depend, in difficult to determine ways, on the overall true market prices of the relevant good. Returning to the example of New York flat, even the Australian with a fear of flying would plausibly still offer a substantial amount of money for the apartment because she could sell it for a profit. In this case, there is some willingness to pay for a good that would confer no consumption utility on the purchaser, and hence the willingness to pay is much higher than would be warranted by the individual's desire to own the apartment. In the case of stimuli such as water bottles or coffee mugs, in contrast, the likely profit that could be made by selling the item is small or zero in the light of transaction costs and hence there is less incentive to provide a WTP that reflects anything other than the small amount of consumption utility the object might confer.

2.13 References

Ashby, N. J. S., Dickert, S., & Glöckner, A. (2012). Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making*, *7*(3), 254–267.

Ashby, N. J. S., Walasek, L., & Glöckner, A. (2015). The effect of consumer ratings and attentional allocation on product valuations. *Judgment and Decision Making*, *10*(2), 172–184.

Bartling, B., Engl, F., & Weber, R. A. (2015). Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. *Journal of the Economic Science Association*, 2015, *1*(1), 72-85.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring Utility by a Single-Response Sequential Method. *Systems Research and Behavioral Science*, *9*(3), 226.

Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, *15*(2), 139.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*(1), 48.

Brown, T. C., & Gregory, R. (1999). Why the WTA–WTP disparity matters. *Ecological Economics*, *28*(3), 323-335.

Brown, T. C. (2005). Loss aversion without the endowment effect, and other explanations for the WTA–WTP disparity. *Journal of Economic Behavior & Organization*, *57*(3), 367-379.

Carmon, Z., & Ariely, D. (2000). Focusing on the forgone: How value can appear so different to buyers and sellers. *Journal of Consumer Research*, *27*(3), 360–370.

Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior*, *25*(4), 439–459.

Ericson, K., & Fuster, A. (2014). The Endowment Effect. *Annual Review of Economics*, *6*(1), 555-579.

Fehr, D., Hakimov, R., & Kübler, D. (2015). The willingness to pay–willingness to accept gap: A failed Replication of Plott and Zeiler, *European Economic Review*, *78*, 120-128.

Fudenberg, D., Levine, D. K., & Maniadis, Z. (2012). On the robustness of anchoring effects in WTP and WTA experiments. *American Economic Journal: Microeconomics*, *4*(2), 131-45.

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

Horowitz, J. K., & McConnell, K. E. (2002). A Review of WTA / WTP Studies. *Journal of Environmental Economics and Management*, *44*(3), 426-447.

Huck, S., Kirchsteiger, G., & Oechssler, J. (2005). Learning to like what you have–explaining the endowment effect. *The Economic Journal*, *115*(505), 689-702.

Isoni, A. (2011). The willingness-to-accept/willingness-to-pay disparity in repeated markets: loss aversion or "bad-deal" aversion? *Theory and Decision*, *71*(3), 409–430.

Isoni, A., Loomes, G., & Sugden, R. (2011). The Willingness to Pay—Willingness to Accept Gap, the" Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment. *American Economic Review*, *101*(2), 991-1011.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 461–74.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*(6), 1325-1348.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, *5*(1), 193–206.

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics*, *99*(3), 507-521.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review*, *79*(5), 1277-1284.

Korobkin, R. (2003). The endowment effect and legal analysis. *Northwestern University Law Review*, *97*(3), 1227-1891.

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

Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of Personality Assessment, 47*(1), 66-75.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, *104*(1), 277-90.

Mazar, N., Kőszegi, B., & Ariely, D. (2014). True context-dependent preferences? The causes of market-dependent valuations. *Journal of Behavioral Decision Making*, *27*(3), 200-208.

Morewedge, C. K., & Giblin, C. E. (2015). Explanations of the endowment effect: an integrative review. *Trends in Cognitive Sciences*, *19*(6), 339–348.

Morewedge, C. K., Shu, L. L., Gilbert, D. T., & Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology*, *45*(4), 947–951.

Nayakankuppam, D., & Mishra, H. (2005). The endowment effect: Rose-tinted and dark-tinted glasses. *Journal of Consumer Research, 32*(3), 390-395.

Pachur, T., & Scheibehenne, B. (2012). Constructing preference from experience: The endowment effect reflected in external information search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1108–1116.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology*, *7*(1), 84–107.

Plott, C. R., & Zeiler, K. (2005). The Willingness to Pay–Willingness to Accept Gap, the. *The American Economic Review*, *95*(3), 530-545.

Reb, J., & Connolly, T. (2007). Possession, feelings of ownership, and the endowment effect possession, feelings of ownership and the endowment effect. *Judgment and Decision Making*, *2*(2), 107–114.

Sharpe, W., Goldstein, D. G., & Blythe, P. (2000). T*he distribution builder: A tool for inferring investor preferences.* Available online at: http://www.stanford.edu/˜wfsharpe/art/qpaper/qpaper.pdf.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1-26.

Strahilevitz, M. A., & Loewenstein, G. (1998). The Effect of Ownership History on the Valuation of Objects. *Journal of Consumer Research*, *25*(3), 276-289.

Simonson, I., & Drolet, A. (2004). Anchoring Effects on Consumers' Willingness-to-Pay and Willingness-to-Accept. *Journal of Consumer Research*, *31*(3), 681-690.

Sugden, R., Zheng, J., & Zizzo, D. J. (2013). Not all anchors are created equal. *Journal of Economic Psychology*, *39*, 21-31.

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

Trautmann, S. T., Vieider, F. M., & Wakker, P. P. (2008). Causes of ambiguity aversion: Known versus unknown preferences. *Journal of Risk and Uncertainty*, *36*(3), 225–243.

Trautmann, S. T., Vieider, F. M., & Wakker, P. P. (2011). Preference Reversals for Ambiguity Aversion. *Management Science*, 57, 1320–1333.

Trautmann, S. T., & Schmidt, U. (2012). Pricing risk and ambiguity: The effect of perspective taking. *The Quarterly Journal of Experimental Psychology*, *65*(1), 195-205.

Tunçel, T., & Hammitt, J. K. (2014). A new meta-analysis on the WTP/WTA disparity. *Journal of Environmental Economics and Management*, *68*(1), 175–187.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061.

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

Walasek, L., Rakow, T., & Matthews, W. J. (2017). When does construction enhance product value? Investigating the combined effects of object assembly and ownership on valuation. *Journal of Behavioral Decision Making*, *30*(2), 144-156.

Weaver, R., & Frederick, S. (2012). A reference price theory of the endowment effect. *Journal of Marketing Research*, *49*(5), 696–707.

Willig, R. D. (1976). Consumer's surplus without apology. *The American Economic Review*, *66*(4), 589-597.

Yang, Y., Vosgerau, J., & Loewenstein, G. (2013). Framing influences willingness to pay but not willingness to accept. *Journal of Marketing Research*, *50*(6), 725–738.

Yechiam, E., & Hochman, G. (2013). Loss attention in a dual-task setting. *Psychological Science*, *25*(2), 494-502.

Yechiam, E., Abofol, T., & Pachur, T. (2017a). The seller's sense: Buying-selling perspective affects the sensitivity to expected value differences. *Journal of Behavioral Decision Making*, *30*(2), 197–208.

Yechiam, E., Ashby, N. J. S., & Pachur, T. (2017b). Who's biased? A meta-analysis of buyer–seller differences in the pricing of lotteries. *Psychological Bulletin*, *143*(5), 543–563.

2.14 Appendices

2.14.1 Appendix I

1) **Experiment 2-Instructions for Buyers**

Participants came to the lab, were greeted and told they were going to be given a flat fee of £7 in total for their participation. They were also told that, depending on their choices, they could earn between £0.00 and £20.00 extra. They completed a filler task (randomness production task) and then took part in our computer based study where they answered questions about the filler task and then they proceeded to the endowment effect study.

We placed a water bottle on each desk and participants were told:

"Have a look at the water bottle that the experimenter put on your desk. We will give you an opportunity to buy this water bottle if you want to. You have the option of either buying the bottle and taking it home with you or keeping all your money. We would like to know the highest amount of money you would be willing to pay for the water bottle. Click the button to proceed for further instructions."

"On the next screen, we will ask you the maximum price you would be willing to pay for the bottle. We will use a random price method to decide whether or not you actually do buy the bottle. At the end of the experiment, the computer will generate a random price for the water bottle. If your buying offer is higher than this price, then you must buy the water bottle for the amount chosen by the computer. If your buying offer is lower than the randomly generated price, then you keep your money and do not get the bottle. On the next screen, you will find some examples."

Next the BDM was explained to them, and the random price method and we also gave some examples.

Examples of the BDM procedure shown in the buyer condition.

Example 1: You indicate that you are willing to pay X for the water bottle, and a higher price of Y is generated. Since the price of Y is greater than your stated buying price of X, you will not buy the bottle and will keep all your money.

Example 2: You indicate that you are willing to pay X for the water bottle, and a lower price of Z is generated. Since the price of Z is less that your stated buying price of X, you will buy the bottle for Z.

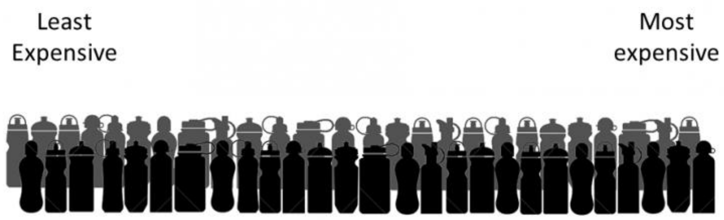
Choose your buying price carefully on the next screen because you will not be able to change it. Because the price is chosen by the computer at random, it is in your best interest to state your true buying price.

"If you have any questions, please raise your hand and an experimenter will come to assist you."

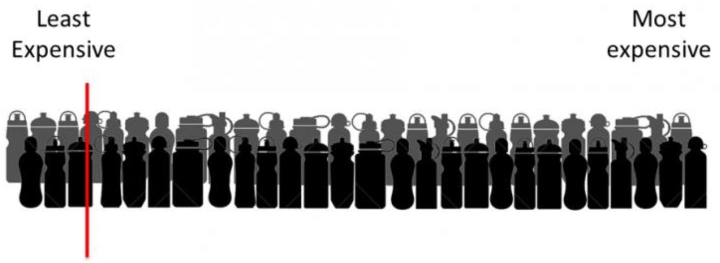
In the box below please enter the highest amount of money, in pounds, that you would be willing to pay for the water bottle (you can use up to two decimal points):

Next, we asked participants some questions about consumer goods.

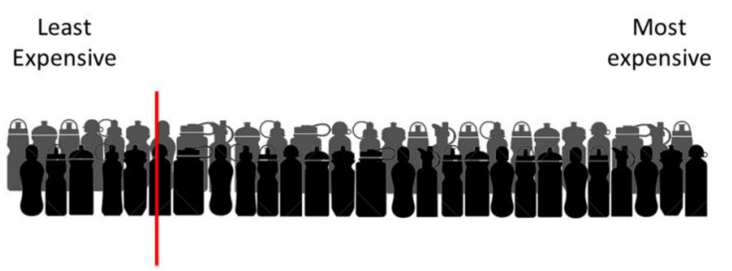
In the following part of the study you will be asked some questions about consumer goods. Please read the following instructions carefully. On the next screen, we will ask you how much water bottles typically cost. Water bottles are represented below in order of price starting from the cheapest (left) to the most expensive (right) bottle. Of course, you will not know the exact answers, but please give your best estimate. We will compare your responses to real price data for water bottles. The three individuals who give the best estimates for prices of water bottles will be awarded bonus payments of 15, 10 and 5 pounds for the first, second and third place respectively.



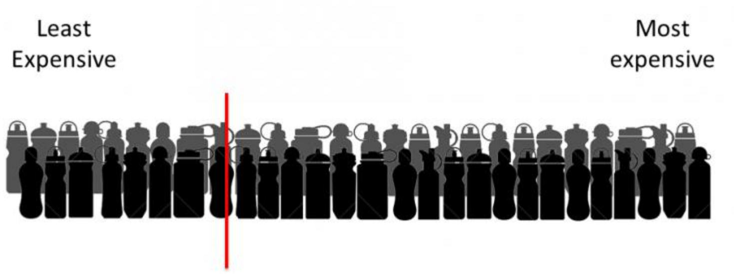
We gave participants nine elicitation questions in random order that related to the picture of black water bottles of different shapes in a row, with Least Expensive on the left of the image and Most Expensive on the right. Each of the nine questions were of the following type:



The line indicates a price. 10% of all water bottles cost less than the price indicated by the line, and 90% cost more. What is the price indicated by the line (in British pounds)?

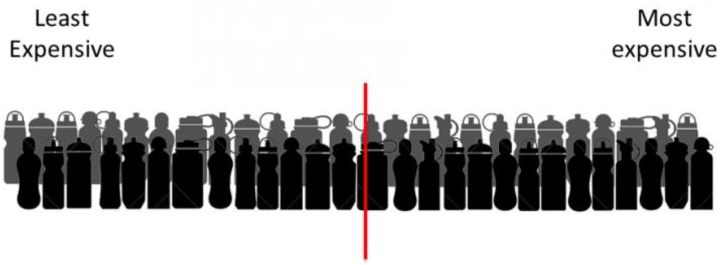


The line indicates a price. 20% of all water bottles cost less than the price indicated by the line, and 80% cost more. What is the price indicated by the line (in British pounds)?

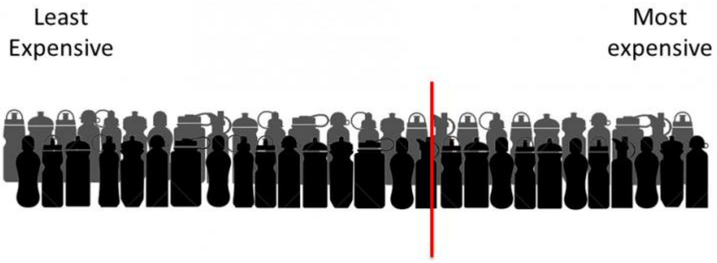


The line indicates a price. 30% of all water bottles cost less than the price indicated by the line, and 70% cost more. What is the price indicated by the line (in British pounds)?

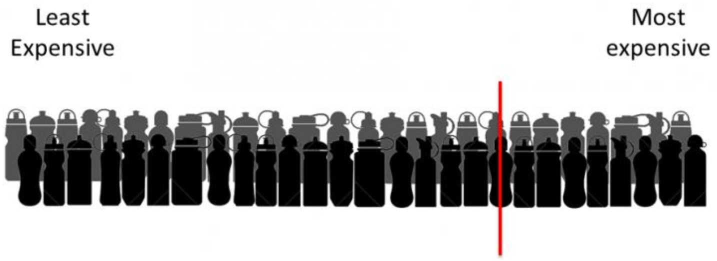
The line indicates a price. 40% of all water bottles cost less than the price indicated by the line, and 60% cost more. What is the price indicated by the line (in British pounds)?



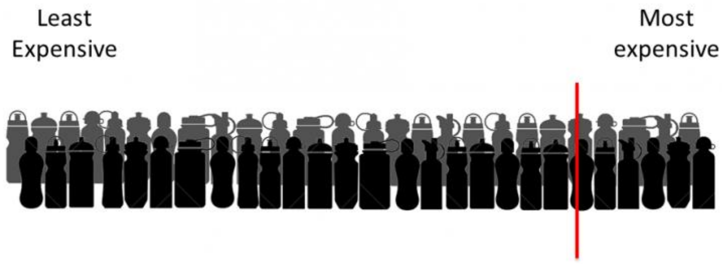
The line indicates a price. 50% of all water bottles cost less than the price indicated by the line, and 50% cost more. What is the price indicated by the line (in British pounds)?



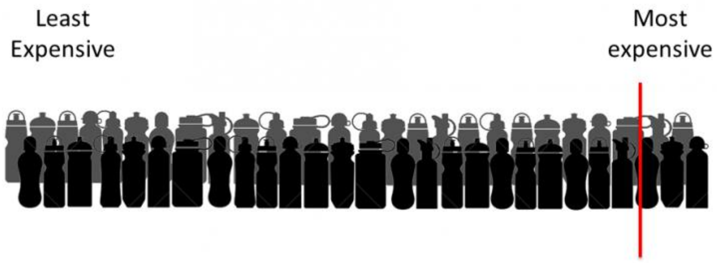
The line indicates a price. 60% of all water bottles cost less than the price indicated by the line, and 40% cost more. What is the price indicated by the line (in British pounds)?



The line indicates a price. 70% of all water bottles cost less than the price indicated by the line, and 30% cost more. What is the price indicated by the line (in British pounds)?



The line indicates a price. 80% of all water bottles cost less than the price indicated by the line, and 20% cost more. What is the price indicated by the line (in British pounds)?



The line indicates a price. 90% of all water bottles cost less than the price indicated by the line, and 10% cost more. What is the price indicated by the line (in British pounds)?

Thank you for completing the consumer goods part of the study. We will compare your responses to real price data for water bottles. The 3 most accurate participants throughout our sessions will receive 20, 15 and 10 pounds respectively.

In a subsequent screen participants were asked:

What do you think the actual market price of the water bottle you were given is?

Please specify your best estimate in the box below in pounds sterling.

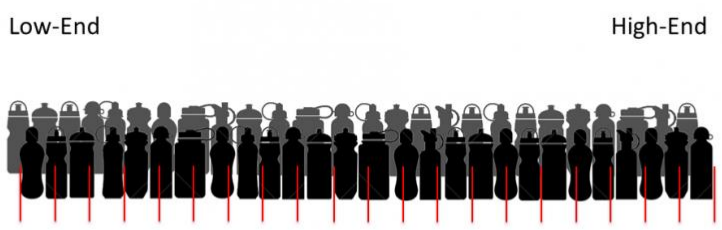
How confident are you of the estimate you specified above? You may not be completely certain about the market price of the water bottle. We would now like you to provide a range such that you are 90% sure that the market price of the water bottle falls within this range.

I am 90% confident that the market price would be more than:

I am 90% confident that the market price would be less than:

In the next screen participants were shown the following:

In the picture below, water bottles are represented in order of quality starting from the most low-end (left) to the most high-end (right). At the high-end water bottles have the most features and best materials and at the low end they have the fewest features and poorest materials. Please indicate by clicking on the appropriate region of the scale below where you believe the water bottle that you have been given ranks in terms of quality. Of course, you will not know the exact answer, but please give your best estimate.



In the previous question, you indicated where the water bottle you were given ranks in terms of quality. The rank you gave is represented by a green rectangle on the graphic below. You may not be completely certain about where the water bottle ranks in terms of quality. We would now like you to provide a range such that you are 90% sure that the quality of the water bottle falls within this range. Please indicate on the first graphic your LOWER limit for the range, so that you are 90% sure that the quality of the water bottle is HIGHER than this limit. You may click on the green rectangle if you believe that the LOWER limit is included in that region. On the second graphic, we would like you to indicate your HIGHER limit for the range, so that you are 90% sure that the quality of the water bottle is LOWER than this limit. You may click on the green rectangle if you believe that the UPPER limit is included in that region. If you are completely certain about where the water bottle ranks in terms of quality, please just click on the green rectangle in both graphics.

Please indicate your LOWER limit for the quality of the water bottle.

Please indicate your HIGHER limit for the quality of the water bottle.

We previously asked you to state the highest amount of money you would be willing to pay for the water bottle. Remember that the computer will randomly generate a price and compare it with your offer. If the price is lower than your stated offer, you will buy the water bottle. If it is higher than your offer you will not buy the water bottle.

If the random price generated was bigger than a buyer's WTP then the participant saw:

Please do not progress from this screen until you have spoken to an experimenter. Your maximum offer was {ChoiceTextEntryValue} and the price generated was {Generated Price}, which is greater than your price. Therefore, you will not buy the item. Please now indicate by raising your arm that you have completed the study. An experimenter will come to give you a form so that you can write down the maximum offer you indicated, the generated price and the outcome of your transaction. After you have completed the form, present it to the experimenter in order to finalise your transaction.

If the random price generated was smaller than a buyer's WTP then the participant saw:

Please do not progress from this screen until you have spoken to an experimenter. Your maximum offer was {ChoiceTextEntryValue} and the price generated was {Generated Price}, which is less than your price. Therefore, you will buy the item for {Generated Price}. Please now indicate by raising your arm that you have completed the study. An experimenter will come to give you a form so that you can write down the maximum offer you indicated, the generated price and the outcome of your transaction. After you have completed the form, present it to the experimenter in order to finalise your transaction.

2) **Experiment 2-Instructions for Sellers**

Participants came to the lab, were greeted and told they were going to be given a flat fee of £3 in total for their participation. They were also told that, depending on their choices, they could earn between £0.00 and £20.00 extra. They completed a filler task (randomness production task) and then took part in our computer based study where they answered questions about the filler task and then they proceeded to the endowment effect study. The sellers experiment was the same, except that sellers were given the water bottle immediately when we welcomed them into the experiment and told it was theirs to keep.

After the filler task participants were told the following:

"At the beginning of this experiment we gave you a water bottle. This bottle is yours to keep if you want. However, we will give you an opportunity to sell your water bottle if you want to. You have the option of either selling your bottle that you were given or keeping it and taking it away home with you. We would like to know the lowest amount of money you would be willing to sell the water bottle for. Click the button to proceed for further instructions. "

"On the next screen, we will ask you the minimum price you would be willing to sell your bottle for. We will use a random price method to decide whether or not you actually do sell the bottle. At the end of the experiment, the computer will generate a random offer for the water bottle. If your selling price is lower than this offer, then you must sell the water bottle for the amount offered by the computer and we will give you this amount in cash. If your selling price is higher than the randomly generated offer price, then you must keep the bottle. On the next screen, you will find some examples."

Examples of the BDM procedure shown in the seller condition.

Example 1: You indicate that you are willing to sell the water bottle for X, and a higher offer price of Y is generated. Since the offer price of Y is greater than your stated selling price of X, you will sell the bottle and receive Y.

Example 2: You indicate that you are willing to sell the water bottle for X, and a lower offer price of Z is generated. Since the offer price of Z is less that your stated selling price of X, you will keep the bottle.

"Choose your selling price carefully on the next screen because you will not be able to change it. Because the offer is chosen by the computer at random, it is in your best interest to state your true selling price.

If you have any questions, please raise your hand and an experimenter will come to assist you."

"In the box below please enter the lowest amount of money, in pounds, that you would be willing to sell the water bottle for (you can use up to two decimal points):"

The remaining parts of the experiment were the same for sellers as they were for buyers.

At the end of the experiment participants were reminded:

"We previously asked you to state the lowest amount of money you would be willing to accept for the water bottle. Remember that the computer will randomly generate an offer and compare it with your selling price. If the offer is lower than your stated selling price, you will keep the water bottle. If it is higher than your selling price you will sell the water bottle for the randomly generated offer.

If the random price generated was bigger than a seller's WTA, then the participant saw:

Please do not progress from this screen until you have spoken to an experimenter. Your minimum price was {ChoiceTextEntryValue} and the offer price generated was {Generated Price}, which is greater than your price. Therefore, you will sell the item for {Generated Price}. Please now indicate by raising your arm that you have completed the study. An experimenter will come to give you a form so that you can write down the minimum price you indicated, the generated offer price and the outcome of your transaction. After you have completed the form, present it to the experimenter in order to finalise your transaction.

If the random price generated was smaller than a seller's WTA, then the participant saw:

Please do not progress from this screen until you have spoken to an experimenter. Your minimum price was {ChoiceTextEntryValue} and the offer price generated was {Generated Price}, which is less than your price. Therefore, you will keep the item. Please now indicate by raising your arm that you have completed the study. An experimenter will come to give you a form so that you can write down the minimum price you indicated, the generated offer price and the outcome of your transaction. After you have completed the form, present it to the experimenter in order to finalise your transaction.

3) **Experiment 1-Instructions for Buyers**

Instructions were the same as the Instructions for Buyers in Experiment 2 with the exclusion of the elicitations of the quality rank and the market price of the specific water bottle.

4) **Experiment 1-Instructions for Sellers**

Instructions were the same as the Instructions for Sellers in Experiment 2 with the exclusion of the elicitations of the quality rank and the market price of the specific water bottle.

2.14.2 Appendix II

Lognormal vs. normal distribution fitting

In the main section of the manuscript, we fitted participants' elicited price distribution only to a lognormal distribution. This was implemented because prices for water bottles are positive numbers and fitting a normal distribution results in a non-zero probability for a water bottle to have a negative price, which leads to negative quality prices. Despite this, it may be possible for a normal distribution to fit participants' elicited price distributions well. The results show that there are no participants who are substantially (R^2>0.1) better fitted by a normal distribution than a lognormal and that the resulting rank estimates of WTA(P) are similar and exhibit the same patterns as those found by using the lognormal distribution exclusively. In the following sections, we fit participants' data to both normal and lognormal distributions and choose the best fitting distribution for each participant.

Experiment 1

We find data from 41 individuals to be best fit by the lognormal distribution and from 36 to be best fit by the normal distribution. A histogram of the difference in R^2 between these two fits is shown in Figure S1 below.

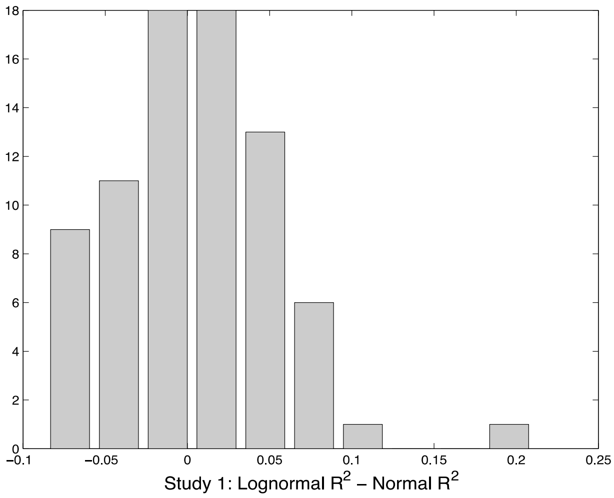


Figure S1. Histogram of the difference in R^2 between normal and lognormal fits in Experiment 1.

The aggregate statistics for the WTA(P) ranks are summarized in Table S1.

Table S1. Aggregate statistics for the WTA(P) ranks in Experiment 1.

	Average Rank	Std. Dev.	Median
Buyers	0.1100	0.1392	0.0683
Sellers	0.3329	0.2457	0.3349

The two values were judged significantly different by both a two-tailed t-test (t (75) = -4.9743, p = 4.0424e-06, ci = [-0.3122, -0.1337]), and a Mann-Whitney U test (Z = -4.4564, p = 8.3347e-06, effect size r = 0.5146).

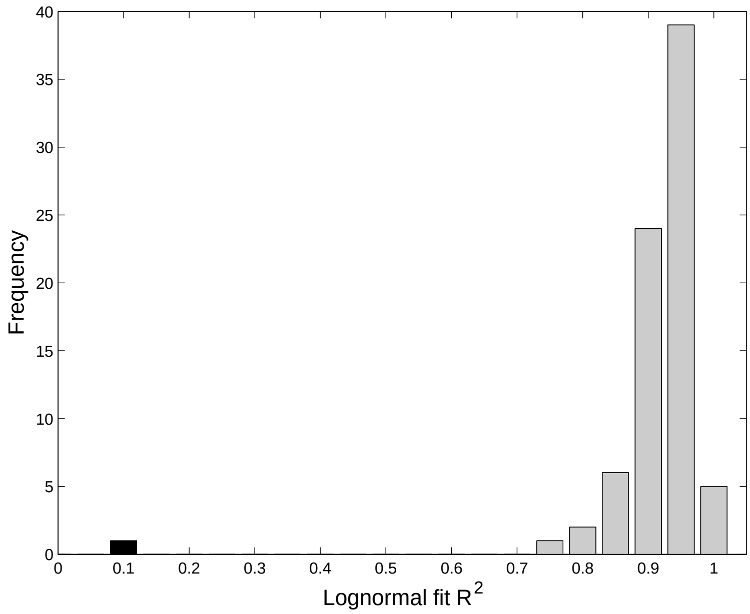


Figure S2. Histogram for the R^2 values of the lognormal fitted distributions to participants' elicited market price percentiles in Experiment 1. The outlier (participant 67) is shown in black.

Experiment 2

We find data from 47 individuals to be best fit by the lognormal distribution and data from 40 to be best fit by the normal distribution. A histogram of the difference in R^2 between these two fits is shown in Figure S3 below, where we see that overall the lognormal distribution fits better and at worst it has an R^2 that is lower by less than 0.1.

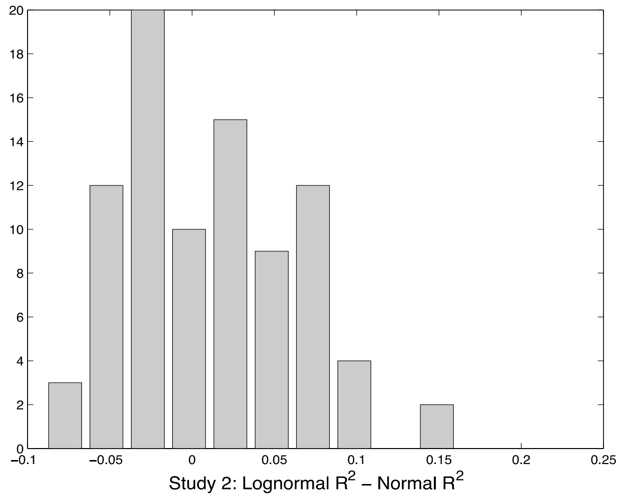


Figure S3. Histogram of the difference in R^2 between these two fits in Experiment 2.

The aggregate statistics for the WTA(P) ranks that come from fitting a normal distribution are summarized in Table S2.

Table S2. Aggregate statistics for the WTA(P) ranks in Experiment 2.

	Average Rank	Std. Dev.	Median
Buyers	0.1744	0.1915	0.1216
Sellers	0.3264	0.2639	0.2240

The two values were judged significantly different by both a two-tailed t-test: (t (85) = -3.0690, p = 0.0029, ci = [-0.2505, -0.0535]), and a Mann-Whitney U test (Z = -2.9884, p = 0.0028, effect size r = 0.3204).

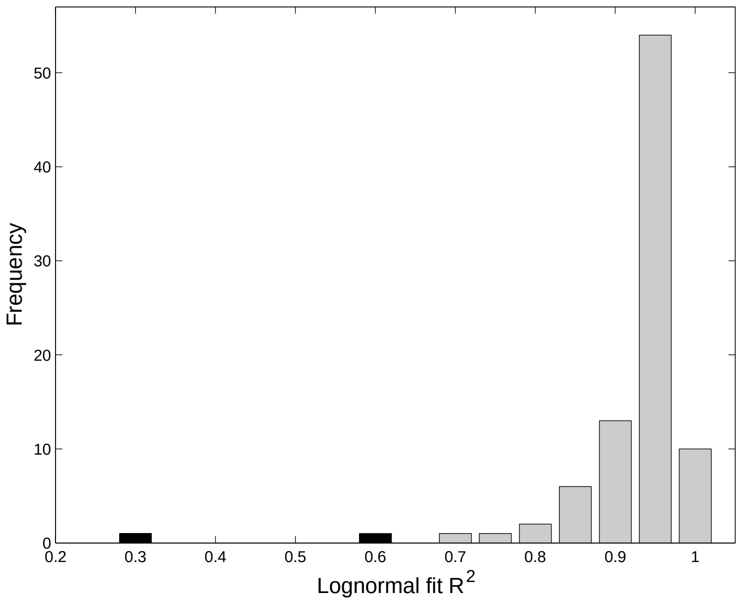


Figure S4. Histogram for the R^2 values of the lognormal fitted distributions to participants' elicited market price percentiles in Experiment 2. Outliers (participants 131 and 135) is shown in black.

2.14.3 Appendix III

Analysis of order effects in Study 2

Results from ANOVAs were conducted to determine if we can combine the sessions where the order of the quality rank questions and market price questions for the specific water bottle were counterbalanced. We had a 2x2 design, with the factors being buy or sell condition and order of quality question and market price question.

Table S3. ANOVA output results (quality rank as the dependent variable).

Source	Sum Sq.	d.f.	Mean Sq.	*F*	*p*
Ownership status	0.08851	1	0.08851	2.81	0.0975
Quality question or Market question first	0.00087	1	0.00087	0.03	0.8689
Interaction term	0.09313	1	0.09313	2.95	0.0893
Error	2.77646	88	0.03155		
Total	2.94457	91			

There is not a strong difference in quality ranks between buyers and sellers (p = 0.0975). There is also a mild interaction effect (p values above 0.05). We can see this by looking at the means in each of the four groups in Table S4.

Table S4. Means of quality ranks in each of the groups.

	Buyer	Seller
Quality First	0.2631	0.3893
Market Price First	0.3208	0.3192

Table S5. ANOVA output results (quality rank range as the dependent variable).

Source	Sum Sq.	d.f.	Mean Sq.	*F*	*p*
Ownership status	0.02924	1	0.02924	6.33	0.0137
Quality question or Market question first	0.00943	1	0.00943	2.04	0.1566
Interaction term	0.01547	1	0.01547	3.35	0.0706
Error	0.40636	88	0.00462		
Total	0.45733	91			

We see a broadly similar pattern to that seen in Table S3. There is a clear difference in quality range between buyers and sellers (p = 0.0137) but there is a weak interaction effect and a weak effect of the ordering of the two sections (quality rank question for the specific water bottle appearing first or market price question first). We can see this by looking at the means in each of the four groups in Table S6.

Table S6. Means of quality rank ranges in each of the groups.

	Buyer	Seller
Quality First	0.0801	0.1420
Market Price First	0.1265	0.1363

Table S7. ANOVA output results (Market Price as the dependent variable).

Source	Sum Sq.	d.f.	Mean Sq.	*F*	*p*
Ownership status	2.483	1	2.4829	0.28	0.5977
Quality question or Market question first	16.314	1	16.3138	1.84	0.178
Interaction term	8.783	1	8.7833	0.99	0.3219
Error	778.836	88	8.8504		
Total	805.478	91			

None of the treatments had any effect on the market price.

Table S8. ANOVA Output results (Market Price Range as the dependent variable).

Source	Sum Sq.	d.f.	Mean Sq.	*F*	*p*
Ownership Status	99.88	1	99.8764	3.3	0.0727
Quality question or Market question first	7.48	1	7.48	0.25	0.6202
Interaction term	10.04	1	10.0414	0.33	0.566
Error	2540.24	84	30.2409		
Total	2653.34	87			

There is only one weak effect between buyers and sellers. The means of each group can be seen in Table S9 below.

Table S9. Means of market price ranges in each of the groups.

	Buyer	Seller
Quality First	4.9250	7.7386
Market Price First	6.1870	7.6458

2.14.4 Appendix IV

Computation of the rank of WTA/WTP in the quality price distribution

To find the rank of WTA/WTP in the quality price distribution, we require the equation of a straight line that passes through two points (x_1, y_1) and (x_2, y_2), as we use linear interpolation in our calculation. The general equation for a straight line is simply

$$y = mx + b$$

Because the line passes through both points, we have the two conditions:

$$y1 = mx_1 + b$$

$$y2 = mx2 + b$$

Solving the system of simultaneous equations gives $y2 - y1 = m(x2 - x1)$ which allows us to find the slope as

$$m = (y2 - y1)/x2 - x1 = \Delta y/\Delta x$$

To find the value of b, we substitute the value of m back in to the equation with $x = x_1$ and $y = y_1$, and we find the relationship

$$y1 = \frac{(y2 - y1)}{(x2 - x1)} x1 + b$$

Which allows us to solve for b as

$$b = y1 - \frac{(y2 - y1)}{(x2 - x1)} x1$$

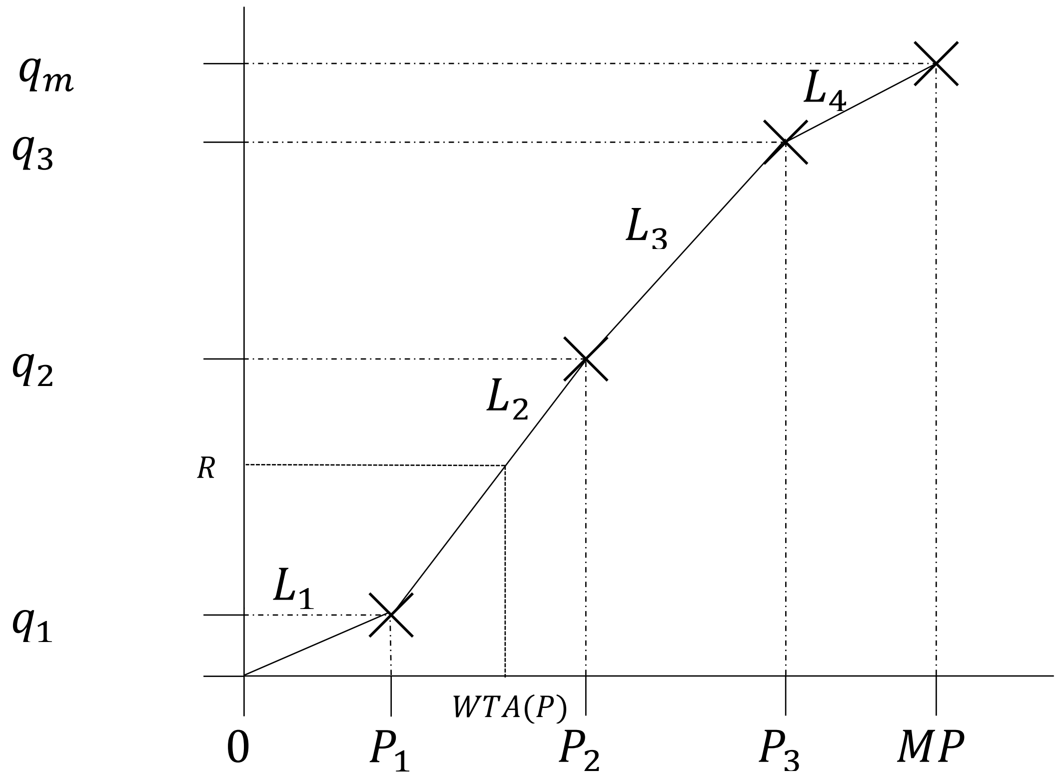
This gives us the equation for the straight line between two points

$$y = \frac{(y2 - y1)}{(x2 - x1)}x + y1 - \frac{(y2 - y1)}{(x2 - x1)}x1$$

Rearranging we find the convenient form

$$y = y1 + \frac{(y2 - y1)}{(x2 - x1)}(x - x1)$$

We use this formula to construct a linear interpolation which allows us to compute the rank of WTA/WTP in the quality price distribution. The scheme for this interpolation is shown by the diagram below. The values for the quality ranks are given as: $q1 = 0.1, q2 = 0.5, q3 = 0.9, qm = 0.99$.



Depending on the value of WTA(P) we use one of four equations to determine the value of the rank, R. The equation we use is determined by the set of inequalities given below; where WTA stands for either WTA or WTP.

If $WTA \leq P_1$

$$R = 0 + (q_1 - 0)\frac{(WTA - 0)}{(P_1 - 0)} = q_1\frac{WTA}{P_1}$$

If $P_1 < WTA \leq P_2$

$$R = q_1 + (q_2 - q_1)\frac{(WTA - P_1)}{(P_2 - P_1)}$$

$P_2 < WTA \leq P_3$

$$R = q_2 + (q_3 - q_2)\frac{(WTA - P_2)}{(P_3 - P_2)}$$

If $P_3 < WTA \leq MP$

$$R = q_3 + (q_m - q_3)\frac{(WTA - P_3)}{(MP - P_3)}$$

As an example computation, suppose the LQP, QP and HQP of the participant were £3, £7 and £10 respectively, and the maximum price, MP, was equal to £15. In this case, the CDF of the APD, which we denote $F(p)$ where p is WTA(P), would be equal to:

$$F(WTA) = \begin{cases} \frac{0.1 \times WTA}{3} & if\ WTA < 3 \\ 0.1 + 0.4 \times \frac{(WTA - 3)}{7 - 3} & if\ 3 \leq WTA < 7 \\ 0.5 + 0.4 \times \frac{(WTA - 7)}{10 - 7} & if\ 7 \leq WTA < 10 \\ 0.9 + 0.09 \times \frac{(WTA - 10)}{15 - 10} & if\ 10 \leq WTA < 15 \\ 1 & if\ WTA \geq 15 \end{cases}$$

2.14.5 Appendix V

Pairwise comparisons for all valuation measures in Study 2

Table S10. Pairwise t-tests.

Comparison	df	sd	*t*	*p*	Confidence Interval
Buyer Quality Seller Quality	85	2.98	-0.64	0.526	[-1.68; 0.86]
Buyer Quality Buyer WTP	42	3.17	5.04	<.001	[1.46; 3.41]
Buyer Quality Seller WTA	85	3.26	0.32	0.746	[-1.16; 1.62]
Buyer Quality Buyer Market	42	2.38	-0.75	0.456	[-1.00; 0.46]
Buyer Quality Seller Market	85	3.08	-1.16	0.250	[-2.08; 0.55]
Seller Quality Buyer WTP	85	2.60	5.09	<.001	[1.73; 3.95]
Seller Quality Seller WTA	43	3.21	1.31	0.198	[-0.34; 1.61]
Seller Quality Buyer Market	85	2.77	0.23	0.820	[-1.05; 1.32]
Seller Quality Seller Market	43	2.73	-0.87	0.389	[-1.19; 0.47]
Buyer WTP Seller WTA	85	2.92	-3.53	<.001	[-3.45; -1.00]
Buyer WTP Buyer Market	42	2.53	-7.01	<.001	[-3.48; -1.93]
Buyer WTP Seller Market	85	2.72	-5.49	<.001	[-4.36; -2.04]
Seller WTA Buyer Market	85	3.07	-0.76	0.451	[-1.81; 0.81]

Comparison	df	sd	*t*	*p*	Confidence Interval
Seller WTA Seller Market	43	2.40	-2.75	0.009	[-1.72; -0.26]
Buyer Market Seller Market	85	2.88	-0.80	0.426	[-1.72; 0.73]

Chapter 3:

Perceptions of Income and Wealth Inequality, Ranks and Subjective Well-Being

“An imbalance between rich and poor is the oldest and most fatal ailment of all republics.”-Plutarch.

3.1 Introduction

It is important to understand the effects of economic inequality on a societal and on an individual level and to determine whether or how it affects citizens' well-being. Nevertheless, it must be the case that subjective perceptions of inequality matter, and one motivation for the research in the present chapter is the possibility that the small or absent effects of inequality in the subjective well-being literature reflect the fact that perceptions of inequality differ from reality, and/or that individual differences exist in those perceptions. In fact, it appears that there is no research connecting perceived inequality to subjective measures of well-being. Moreover, the few studies that have measured people's perceptions of inequality and compared them to objective measures have come to different conclusions.

The aims of this chapter are therefore multiple. We develop a new methodology to allow us to ask (a) whether perceptions of inequality differ from reality, (b) what individual differences predict perceptions of inequality (e.g., personal income/wealth, age, gender and political ideology), and (c) whether subjective inequality might predict subjective well-being. We do this in two separate inequality domains: income inequality (Study 1) and wealth inequality (Study 3) which are often treated as identical in the literature. We also conducted a re-test study (Study 2) in order to investigate how or if our measure of perceived inequality changed over time within individuals as well as to check for changes in other individual measurements between the original (Study 1) and the re-test study (Study 2).

3.2 Economic Inequality

Rising inequality has become one of the defining economic and social issues of the twenty first century. Since the 1970s, economic inequality in the United States has steadily risen (Piketty & Saez, 2003). High inequality is associated with many social ills, such as an increased death rate (Zheng, 2012), reduced social mobility (Kraus & Tan, 2015) decline in health (Pickett & Wilkinson, 2015), lower productivity (Norton, 2014), economic instability (Krugman, 2010) and reduced political equality **(**Solt, 2008). In recent years, economic inequality has become increasingly relevant in British and American political discourse, being cited as one of the critical issues in the

2016 US Presidential Election (Darvas & Efstathiou, 2016), as well as being a possible factor behind the decision of the United Kingdom electorate to leave the European Union (Goodwin, 2016).

3.3 Inequality and its Relationship to Societal and Individual Well-Being

On a societal level, there is an overwhelming amount of evidence that, at least in developed nations, higher economic inequality leads to a plethora of social ills (Wilkinson & Pickett, 2011). For example, there is evidence that higher economic inequality leads to increased death rates (Kennedy 1996, Zheng, 2012); increased homicide rates (Daly, Wilson, & Vasdev, 2001) and general crime rates (Fajnzylber, Lederman, & Norman, 2002); higher levels of infant mortality (Wilkinson & Pickett, 2011); reduced social mobility (Kraus & Tan, 2015); decline in health[13], including increased obesity and calorie consumption levels (Pickett, Kelly, Brunner, Lobstein, & Wilkinson, 2005); economic instability (Krugman, 2010); reduced political equality (Solt, 2008), and decreased levels of trust and cooperation (Kawachi Kennedy, Lochner, & Prothrow-Stith, 1997; Paskov & Dewilde, 2012) as well as increased political polarisation (McCarty, Poole, & Rosenthal, 2006) among others.

It seems natural to suppose that because income inequality has such wide-ranging negative effects on a society that there should be a corresponding effect of income inequality on individual well-being. Indeed, reports of subjective well-being measures have been found to correlate positively and strongly with psychological and economic objective well-being measures (Oswald & Wu, 2010). However, evidence for such an effect is mixed. Alesina, Di Tella and MacCulloch (2004) found a negative relationship between inequality and happiness in European countries, with no corresponding significant result for the US. Within European countries, wealthy individuals' happiness was not affected by income inequality while poor individuals' happiness was negatively affected. Within the US, poor individuals reported higher happiness levels with increasing income inequality, while for wealthy individuals,

[13] Although Pop, Van Ingen & Van Oorschot (2012) and Lynch et.al 2004 found no such effects on population health for highly developed countries but only for low and middle developed countries with the exception of U.S.A.

inequality had no effect. Oishi, Kesebir and Diener (2011) analysed 40 years of longitudinal US data, finding that lower self-reported happiness was associated with increased inequality. Hagerty (2000) studied the relationship between happiness and income inequality at a community level in the United States as well as in 8 other countries over a 25-year period. He specifically investigated the effects of range and skewness of the income distribution, finding a negative relationship between these and happiness. Oshio and Kobayashi (2011) found the same relationship for a randomly chose sample of Japanese districts. Their study also revealed that inequality had the biggest effect on those who did not have stable employment and were already low on happiness rates. Tomes (1986) offered more evidence of individual related differences, showing that income inequality decreased the subjective well-being of females in Canada but increased that of males. Using data from the European and World values surveys from 84 countries, Verme (2011) found that income inequality had a negative effect on life satisfaction, though the relationship was not robust to the introduction of fixed effects of country and year.

Several other studies have found an opposite effect. Clark (2003), using the British Household Panel Survey, found a positive relationship between income inequality and happiness. Berg and Veenhoven (2010) also found a positive relationship between income inequality and a variety of happiness measures in a study of 119 countries. O'Connell (2004), using European Community Household Panel data, found a positive relationship between income inequality and life satisfaction. In developing nations, both Ott (2005) and Helliwell and Huang (2008) found a positive association between subjective well-being and income inequality, with no such effects found for wealthy countries. Senik (2004) found increased happiness among Russian individuals whose reference's group income had increased. However, overall, she did not find an effect of inequality.

There have also been studies finding a negligible relationship between well-being and income inequality. Bjørnskov, Dreher and Fischer (2008) found, by utilising longitudinal data over 5 years across 60 countries, that the skewness of a country's income distribution did not affect the happiness of its citizens. Fahey and Smyth (2004) found that inequalities in European societies had little effect on life satisfaction but the effects were significant within poorer European societies. Grosfeld and Senik

(2010) argued that the effect of inequality on well-being has reversed over time in Poland.

A recent review by Kelley and Evans (2017a) found no effect of income inequality on subjective well-being in developed nations. Their study used data from 68 countries between 1981-2009 from the pooled World Values/European Values Surveys with more than 200,000 respondents. Kelley and Evans (2017b) also found no effect of inequality on happiness for rich nations but for developing nations there was a slight increase attributed to foreshadowing future prosperity. The recent meta-analysis by Ngamaba, Panagioti, and Armitage (2017; included 24 studies between 1980-2017) found that the relationship between income inequality and subjective well-being was very close to zero and not significant and remained nonexistent even if the subjective well-being measure, or the geographic region was changed.

So far there have been very few efforts to understand these mixed results. One explanation proposed by Verme (2011) is the collinearity problems relating to the use and choice of country and year effects. Alesina et al. (2004) and Clark (2003) tried to explain their findings by utilising the idea of the "tunnel effect"[14] (Hirschmann & Rothschild, 1973). They related this concept to people's perceptions about social mobility and anticipation of future earnings which can lead the poor in believing that opportunities exist to increase their incomes even under high inequalities.

Recently, Brown, Boyce and Wood (2018) also addressed this paradox, arguing that a rank-based model of well-being and income predicted an interaction effect of income inequality on the relationship between income and subjective well-being. Indeed, in their inter-country study they found that the relationship between income and well-being was stronger in countries with more equal income distributions. Their approach originated with the notion that the driver of the relationship between income and well-being is comparative, rather than absolute; that rank of income, rather than income itself is important.

To summarise, the results are mixed as to whether inequality affects subjective well-being. Moreover, even if studies find negative relationships between these

[14] There is a traffic jam in a tunnel, although your lane is not moving, you can see the lane next to you slowly moving increasing your belief that you are too not going to be stationary soon. Therefore, paying attention to future progress you can make compared to noticing that others are doing better than you (in traffic terms).

variables, the effects have been small or even absent compared to what one would expect by knowing the degree of the adverse effects of inequality on society. Perhaps, this is because almost all studies have used objective, rather than subjective, measures of inequality.

3.4 Income, Rank of Income and Individual Well-being

Although the effects of inequality on well-being have not been clearly established, effects of income on well-being have been widely documented and hence income is a necessary factor to consider alongside income inequality. Moreover, as with inequality, perceptions of income distributions may influence the relation between income and well-being to the extent that perception of relative, rather than absolute, income determines well-being.

It is well established (Easterlin, 1974) that there is a strong within-country relationship between income and subjective well-being at a point. This relationship persists in more recent studies. For example, Stevenson and Wolfers (2008) found, using 2007 Gallup data, that there was a strong relationship between life satisfaction and income in the United States, with no satiation point even at high incomes ($500,000). Sacks, Stevenson and Wolfers (2010) found similar results.

The relationship between income and subjective well-being is far from straightforward however. Easterlin's paradox (Easterlin, 1974) notes that within a country subjective well-being does not grow over time, despite increases in incomes/GDP over that period. Indeed, there is evidence that the relationship between income and well-being is comparative, rather than absolute; driven by a comparison with a group rather than the absolute level of income itself. For example, Clark and Oswald (1996) found that people derive higher satisfaction from having a high income in comparison with others. Furthermore, Luttmer (2005) found that individuals' happiness increased as their income increased in comparison with their neighbours'. Brown, Gardner, Oswald and Qian (2008), using both survey data and laboratory experiments, found that the relative rank of a person's wage predicted their satisfaction with their wage. More generally, Boyce, Brown and Moore (2010), using data from the British Household Panel Survey (BHPS), found that the rank of an individual's income, rather than their income itself, better predicted life satisfaction, particularly when the ranks were taken within relevant geographic, gender, education and age reference groups. Similar results have also been found by Clark and Senik (2012) regarding the relationship between well-being and income rank in Chinese villages, as well as Clark, Westergård-Nielsen and Kristensen (2009), who found that income rank within one's neighborhood had a positive effect on subjective well-being.

Despite satisfaction and perceptions, concerns with relative rank have also been shown to affect reward activity in the brain (Fliessbach et al., 2007) ill-health in humans and animals (Sapolsky, 2005), as well as subjective judgments of diverse quantities within a comparison set (Stewart, Chater, & Brown, 2006). It is therefore important to also document its influence on well-being and not just concentrate on the effects of inequality or perceived inequality of income/wealth.

3.5 Perceptions of Inequality –Literature and Methodologies

Many studies have examined the relationship between actual levels of economic inequality in various countries and measures other than subjective well-being, such as education, health and other important societal and individual outcomes. However, individuals' decision-making and choices must necessarily be influenced by their own subjective perceptions of inequality rather than objective inequality. More importantly, while the ills of inequality suggest that it is in the interest of most societies to reduce inequality, society must be able to perceive the inequality (and reductions in it) in order to be influenced by it. Indeed, Gimpelson and Treisman (2015) argue that discussion of the political effects of inequality should be replaced by discussion of the political effects of perceived inequality. In addition, Cansunar (2016) argues that, when making decisions about taxation policies, individuals depend upon their (inaccurate) perceptions of income inequality and their position within a perceived income distribution.

Perhaps then the absent or negligible effects of inequality in the subjective well-being literature discussed previously reflect perceptions of inequality that differ from reality, and/or individual differences in those perceptions. It is therefore important to guide our research efforts towards an accurate elicitation and understanding of these perceptions and their relationship with objective measures. Unfortunately, research has mainly focused on the effects of actual levels of income inequality, with the literature on the perceptions of inequality being new and relatively small.

Studies consistently show, that measures of perceived economic equality differ greatly from objective measured values (Gimpelson & Treisman, 2015; Brunori, 2015; Engelhardt & Wagener, 2014) and indeed vary also between studies of differing methodology. In a controversial paper, Norton and Ariely (2011) (see also Norton, Neal, Govan, Ariely & Holland, 2014 for an Australian study) found that Americans dramatically underestimated (by a factor of more than ten) the true level of wealth inequality when asked what they perceived to be the percentage of wealth held by each wealth quintile (the definition of net worth was used). Underestimation of inequality was also found in when people stated their beliefs about the ratio of pay between CEOs and workers in a cross-national study by Kiatpongsan and Norton (2014). Less dramatic underestimation of wealth inequality was found by Eriksson and Simpson (2012, 2013), who modified the methodology of Norton and Ariely (2011) to elicit average wealth in each quintile, rather than percentage of total wealth in each quintile, and who argued that the dramatic underestimation of wealth inequality was due, in part, to methodological issues. Questions that ask for a percentage of total wealth may be more difficult than questions about average wealth. Moreover, if people have negative wealth, which some do, asking for a percentage of the total wealth the first quintile has would make people give positive values as percentages must be positive, while it could be for example that 25% of people are in debt.

Moderate underestimation of income inequality was found by Page and Goldstein (2016), who elicited a probability mass function (PMF) for the perceived income distribution explicitly from each participant in their study. Nevertheless, such a method could introduce anchoring effects and thus biases in responses and their associated elicited distributions because participants were presented with prespecified income ranges that required them to indicate how much mass of the relevant distribution fell within each interval. In an inter-country study, Engelhardt and Wagener (2014) used data from the International Social Survey Programme (ISSP) to construct a countrywide estimate, rather than estimates on an individual level, of perceived income inequality. In this survey, respondents had to indicate the diagram (income distribution) that best represented the society they live in, among five options that ranged between a very unequal society to a very equal society. Responses were aggregated by country and formed a new diagram. Perceived inequality was then measured as the mean income to median income ratio of this diagram. In each country examined, the estimate was below the true value. Using a similar methodology, Cruces, Perez-Truglia, and Tetaz (2013) found that the self-reported distribution of income ranks showed a lower degree of dispersion than the actual distribution.

While the studies above find underestimation of economic inequality, several studies have in contrast shown that individuals overestimate the level of economic inequality. Chambers, Swan and Heesecker (2014) used a variety of methodologies including: eliciting the percentage of individuals with income in three intervals ($0-$35,000, $35,000-$75,000, $50,000+); eliciting the ratio of the average income of the top 20% to the average income of the bottom 20% (the 20-20 ratio); asking forced-choice questions regarding the income of the top 1% and bottom 1%, and eliciting the incomes of the top 20% and bottom 20% directly, as well as eliciting incomes of three high percentiles. Their results showed that individuals consistently overestimated income inequality in the USA, as well as the rise of income inequality over time. They also found a pattern of individuals overestimating the income of high percentile individuals (a pattern that is also observed in the present study). Kuhn (2015) used ISSP data to construct a measure of subjective income inequality from individuals' estimates of the average wage for various professions, finding that, on aggregate, individuals overestimated income inequality. Engelhardt and Wagener (2016) presented Germans with several pre-generated income distribution diagrams, and asked participants which distribution best represented their country, and these authors also found that, on average, participants overestimated income inequality.

A similar methodology was used in a cross-country study by Gimpelson and Treisman (2015) who presented participants with several pre-generated income distribution diagrams with known Gini coefficients, and asked participants which distribution best represented their country. They found that while perceived and actual Gini coefficients were correlated (r = 0.60), there was considerable error, both underestimation and overestimation, on a country-by-country basis. Further country dependence was found by Niehues (2014) who also used ISSP data to compare perceived income distributions to actual distributions and found that in many European countries individuals underestimated the number of people in the middle of the income distribution, whereas Americans underestimated the number of people at the bottom of the income distribution.

In conclusion, the results from the studies on perceptions of inequality are not clear and seem to be methodology dependent.

3.6 Perceived Inequality and Individual Differences

3.6.1 Income

The absent or negligible effects of inequality on subjective well-being may be potentially caused by individual differences that might exist in the perceptions of inequality. Generally, it is assumed that wealthy individuals are opposed to wealth redistribution (e.g. Alesina & Giuliano, 2011; Kaltenthaler, Ceccoli & Gelleny, 2008; Cruces et al., 2013) and that there are rational economic reasons for their doing so**,** with the burden of redistribution falling mainly on the wealthy. Furthermore, in the USA, high income individuals perceive themselves as deserving of their wealth; indeed, Kreidl (2000) finds that high income Americans were more likely to reject structural and embrace individualistic causes of wealth. These results are in line with those of Kaltenthaler et al. (2008) who found that high income Europeans perceived lower inequality (measured as attitudes towards inequality). Perhaps then people's attitudes towards inequality influence their perceptions of how much of it there is. Indeed, Page and Goldstein (2016) found that wealthy individuals overestimated the incomes of the poorest 90% of individuals. Similar results have also been found by Dawtry, Sutton and Sibley (2015) who found that individuals whose social circle was of high income perceived less inequality and Binelli, Loveless and Whitefield (2015) who found that urban based high-income individuals perceive less inequality. Despite this, the evidence is not unanimous. Norton and Ariely (2011) did not find statistically that the estimation of wealth inequality depended on the income of their respondents or other demographic characteristics.

3.6.2 Ideology

Turning now to ideology, Norton and Ariely's 2011 paper found that liberals and conservatives had similar perceptions of wealth inequality, and similar results were also found by Chambers et al. (2014). Kaltenthaler et al. (2008) found that individuals on the left were more likely than those on the right to suggest that eliminating big inequalities in income between citizens is important. Kraus and Tan (2015) found the conservatives were slightly more likely to overestimate social mobility compared to liberals. Also studying social mobility, Davidai and Gilovich (2015) and Chambers et al. (2014) found that conservative individuals perceived more social mobility compared to liberals. Using data from the International Social Survey Programme, Kuhn (2015) showed that those who identify with the political right perceived lower inequality, as measured by the average perceived salary of individuals in various professions. In Page and Goldstein (2016), out of all people with all political leanings, conservatives assigned the highest incomes to poorer individuals.

3.7 Present Work

3.7.1 Aim 1

Among the studies on perceptions of inequality discussed previously some have elicited percentages of wealth or average wealth for quintiles and not the whole distribution (e.g. Norton & Ariely, 2011; Eriksson & Simpson, 2013). Others have elicited the whole distribution of incomes but simultaneously introduced possible anchoring effects (e.g. Page & Goldstein, 2016). Others have elicited average top and bottom incomes for specific percentiles of the distribution (e.g. Chambers et al., 2014). Many of them have used pre-generated income distribution diagrams, and asked participants to choose among these (e.g. Niehues, 2014). None of these methodologies produce results directly comparable to a robust measurement of economic inequality such as the Gini coefficient. Moreover, there is a huge misunderstanding in the literature when it comes to attitudes and perceptions of inequality, with elicited attitudes mistakenly translated to perceptions. Finally, the terms wealth and income inequality have been used interchangeably even within a single paper even when only one of these quantities was actually investigated in a study.

In the present chapter we address these issues by eliciting estimates of the income distribution from participants, as well as estimates of the wealth distribution and hence deriving measures of income and wealth inequality.

Using methods from relative rank theory we develop a new methodology allowing us to explore individuals' perception of income and wealth inequality in a systematic and comprehensive manner. Our methodology is robust, permits self-validation, and allows for the computation and use of more compact summary indices, such as the Gini coefficient. To obtain estimates of perceived income and wealth inequality we elicit estimates of income and wealth percentiles for individuals in the USA. The methodology is founded upon the concept that subjective judgements or estimates of quantities are influenced by the relative ranked position of the quantity within a context (Stewart, Chater & Brown, 2006), with quantities as diverse as fairness (Mellers, 1982) and prices (Niedrich, Sharma, & Wedell, 2001) for example, being determined partly by the relative ranked position they occupy within a comparison context. The distributions are elicited by presenting participants with a graphical representation, a tool that has been found to increase individuals understanding of probability distributions (Goldstein & Rothschild, 2014).

Our methodology also leverages the insights of Eriksson and Simpson (2012, 2013), who argued in favour of eliciting average wealth of individuals in a quintile rather than the percentage of total wealth in each quintile. In a similar manner, our methodology elicits the income and wealth of specific percentiles, rather than the percentage of total income/wealth in each percentile. Our methodology permits a self-consistency check to validate and confirm its robustness. We compute a rank position for each individual within their own elicited income/wealth distribution. By comparing this rank position with an independently elicited subjective rank position (where individuals think they rank in the overall income/wealth distribution) we demonstrate that the elicited distributions accurately portray each individuals' beliefs about the distribution of income/wealth in the USA.

To foreshadow: using this methodology, we find that individuals consistently overestimate the incomes of individuals in each percentile, compared to the Census data for individual incomes. The perceived level of income inequality among our participants, as measured by the Gini coefficient of the elicited aggregate CDF, is only slightly higher than the measured value obtained from the Bureau of Labor Statistics and Census Bureau. In terms of wealth, we find that participants overestimate the wealth of the low percentiles, and of the highest percentile but accurately estimate the middle ones when judged against data from the Survey of Consumer Finance. We also find that the perceived level of wealth inequality among our participants, as measured by the Gini coefficient of the elicited aggregate CDF, was very close to the actual US Gini for wealth.

Using our elicited measures of perceived income and wealth inequality, we are also able to investigate the relationship between individual differences and perceived income and wealth inequality. The principal factors we investigated were personal income, age, gender and political ideology, as measured on a left-right spectrum as well as a conservatism scale that accounts for social and economic conservatism. Our findings indicate that high income/wealth individuals perceive lower levels of income inequality and wealth inequality in the society than low income/wealth individuals. We also show that conservative individuals perceive a lower degree of income/wealth inequality compared to less conservative individuals.

3.7.2 Aim 2

In the well-being literature presented above, most studies investigating the relationship between inequality and well-being use objective measures (actual levels) of inequality. Indeed, as noted by Bjørnskov, Dreher, Fischer, Schnellenbach, and Gehring (2013), it is often assumed that subjective perceptions and objective realities of fairness in society coincide, an assumption that is not necessarily supported by much of the evidence discussed above. In the same study Bjørnskov et al. (2013), using data from the World Values Survey, found that people who perceived the income generating process in their society to be fair had higher self-reported happiness. The issue of perception of fairness was also addressed by Alesina et al. (2004). These ideas are consistent with the notion that perceptions of inequality might affect well-being.

In fact, it appears that there is remarkably little research connecting perceived inequality to subjective measures of well-being. Firstly, we know that not all people attend to or have the same information, so actual levels of inequality may not matter but perceptions might. Secondly, the mixed evidence regarding individual well-being and inequality may reflect individual differences in the perceptions of inequality. If people underestimate inequality for example, their perceptions are lower than reality, any effects of actual levels of inequality on an individual's well-being might be underestimated. The opposite could also be true. Therefore, in this chapter, we also examine the relationship between perceived income and wealth inequality and subjective well-being. This relationship has not been examined so far in the literature.

Moreover, given the evidence discussed above that rank of income matters for an individual's well-being more than income, we could ask: Is this also true compared to perceived inequality? What influences individuals' subjective well-being the most?

To foreshadow: we find, firstly, that an individual's belief about their rank of income/wealth is a better predictor of their subjective well-being than their income/wealth. Indeed, while Norton (2013) claims that 'all ranks are local', it appears that individuals' rank position within their perceived global income/wealth distribution is a better predictor than their personal income/wealth.

While we find no relationship between perceived income/wealth inequality and affect measures, evaluative and aspirational measures of subjective well-being show some very marginal associations. Nevertheless, controlling for an individual's belief regarding their rank of income/wealth diminishes these marginal effects.

3.8 Study 1-Income Inequality

3.8.1 Methodology

3.8.1.1 Design

In this study we developed a new method, based on work on eliciting distributions from other domains. These psychological studies have elicited distributions (inverse CDFs) similar to our account (however not utilizing a visual instrument or random ordering) (e.g. Aldrovandi, Wood, Maltby, & Brown, 2015; Melrose, Brown, & Wood 2012) and showed how people's perception of their relative ranked position in this perceived distribution predicts subjective judgements of quantities varying from mental distress to attitudes towards debt, perceptions of alcoholism and willingness to pay for healthy vs unhealthy foods.

This elicitation allowed us to collect more detailed data about individuals' perceived income distributions than previous studies. An additional feature of our methodology was the elicitation of incomes for the tails of the income distribution – the incomes of the top and bottom 1% of the population. These individuals, particularly the top 1%, have been the subject of much media attention in recent years, and this attention may have influenced participants' beliefs due to availability bias.

3.8.1.2 *Participants*

The experiment was conducted using the Amazon Mechanical Turk platform. Our sample consisted of 1003 participants subject to a United States location filter (M_{age} = 36.6, 47.3% female). The experiment lasted on average 30 minutes and participants were paid $1 for completion.

3.8.1.3 *Procedure*

Prior to the elicitation of incomes participants were given instructions and a graphical representation of what would follow and what would be asked from them.

Participants were told:

> *"Look at the graphic below. Imagine that it represents the entire population of the USA. All residents are ordered from the poorest (left side of the graphic) to the richest (right side of the graphic). In other words, the leftmost person in the row has the lowest income, whereas the rightmost person has the highest income.*
>
> *On the next screen, we will ask you some questions about the incomes of other people."*

The income distributions were elicited by asking participants for their estimate of the incomes of 11 different percentiles. An example of such a question is shown below:

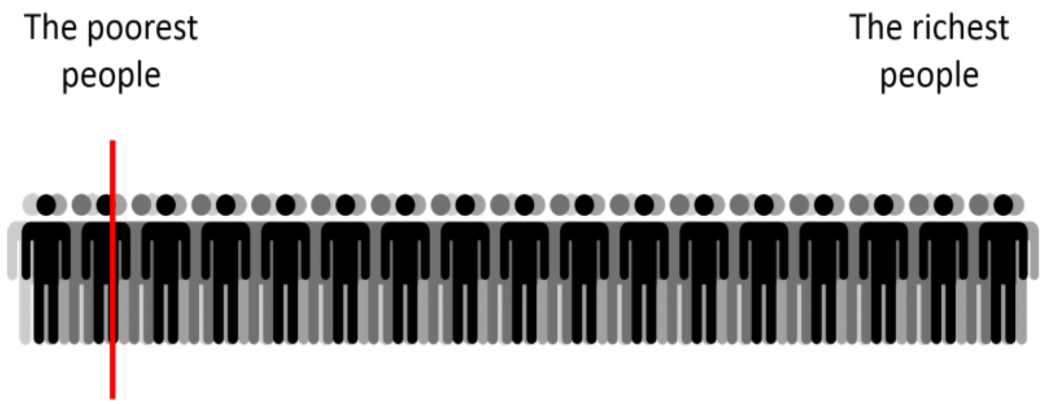


Figure 3.1. Example question shown to participants in Study 1.

"The red line points towards those whose income is higher than 10% of the US population, and lower than 90% of the US population.

What do you think is the income of such a person (per year)?"

Participants saw a series of these pictures for the different percentiles 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 99. The order of the questions was randomised for each participant. The elicited data were therefore in the form of an inverse cumulative distribution function for each participant. From these data we were then able to compute each participant's perception of inequality in the form of the Gini coefficient as well as calculating the rank position of their own personal annual income in their perceived income distribution (inferred rank). We also elicited subjective measures of the participant's income rank by asking them directly to choose into which percentile of the income distribution they fell (subjective rank).

Participants were asked:

"Think again about your own income, i.e. how much you earn per year (wages, salary, commissions, bonuses, tips).

Where do you think you rank in terms of income (per year) comparing with the rest of the US population? Click on the picture to indicate your approximate position."

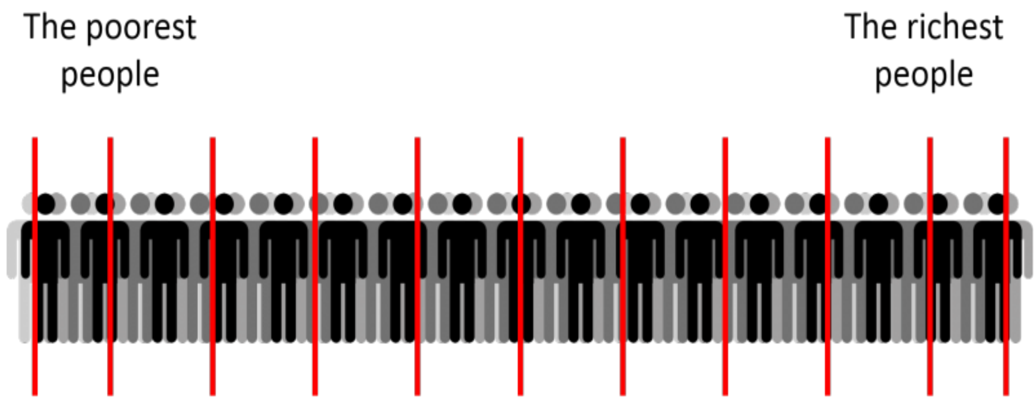


Figure 3.2. Subjective Rank question.

This allowed us to validate our methodology by examining the degree to which the two different measures of rank, inferred and subjective, coincided.

We also elicited information about subjects' well-being as well as their political ideology. These measurements allowed us to test the relationship between perceived inequality, rank, subjective well-being and political ideology. Political ideology was measured using the Everett scale (Everett, 2013) as well as a left to right scale ranging from very liberal to very conservative ("Please select which of the following best represents your views on politics today, where 1 represents extremely liberal and 7 represents extremely conservative"). We also elicited political affiliation (Democrat, Republican, Independent and Other).

Six measures of subjective well-being measured on 0-10 scale were elicited, including overall life satisfaction ("How satisfied are you with your life nowadays?"; used in the UK Office of National Statistics -ONS-, Annual Population Survey, 2012), happiness ("Overall, how happy did you feel yesterday?"; used in ONS, 2012 survey), anxiety on the previous day ("Overall, how anxious did you feel yesterday?"; used in ONS, 2012 survey), eudemonic well-being ("Overall, to what extent do you feel that the things you do in your life are worthwhile?"; used in ONS, 2012 survey) and two ladder estimates (Candril ladders used in Bjørnskov, 2010). Ladder 1 ("Look at the ladder presented below. Imagine that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?"), ladder 2 ("On which step do you think you will stand about five years from now?"). These measures were obtained at the beginning of the study as per the OECD guidelines on the investigation of subjective well-being (OECD, 2013) because introducing the experimental tasks first could have influenced responses on the well-being questions. We included this set of measures also known as "core measures" which is the minimal set of life evaluation and affect measures of subjective well-being shown to be valid and comparable by OECD (2013).

Other questions asked in the beginning of the study included demographic information of age, gender annual personal income ("What is the total amount that you have earned in the last 12 months. This should include all wages, salary, commissions, bonuses or tips from all jobs in $"), and the CEO-worker index ("We would like to know what you think people in particular jobs actually earn. Please write in how much you think they usually earn each year, before taxes. Many people are not exactly sure about this, but your best guess will be close enough. This may be difficult, but it is very important. So please try. How much do you think a chairman of a large national corporation earns per year? How much do you think an unskilled worker in a factory earns per year?"). This allowed us to obtain additional measures of income inequality, to compare with the Gini coefficients computed for each individual from the elicited income distributions.

3.9 Results

3.9.1 Measuring Aggregate Perceptions of Income Inequality

The study was completed by 1003 participants. 279 individuals who stated that they had a yearly income less than $10,000 were excluded from all the analysis presented in this section[15]. This was due to the assumption that it would make little sense to examine the relationship between income and subjective well-being for individuals who either have no income or are likely to be surviving on supplements (e.g. food stamps). Moreover, these participants might not be similar to other income earners, representing part time employees or being unemployed. In the US the minimum wage is approximately $7 per hour, with individuals working part time earning around $12,000 per year. Therefore, we decided to exclude individuals with an income less than $10,000 as they might not have a job or might be working part time.

Figure 3.3 shows the aggregate income distribution elicited from participants, computed as the median income in each percentile, along with the personal income distribution for 2015 taken from the Bureau of Labor Statistics and Census Bureau Current Population Survey (CPS)[16] (personal income data which were given in intervals). The median was used rather than the mean as participants' responses for each income percentile demonstrated high positive skew.

[15] The analysis was also conducted with all participants included. The results obtained in this way did not differ qualitatively from those reported here and later in this chapter.

[16] Table PINC-01, http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-01.html.

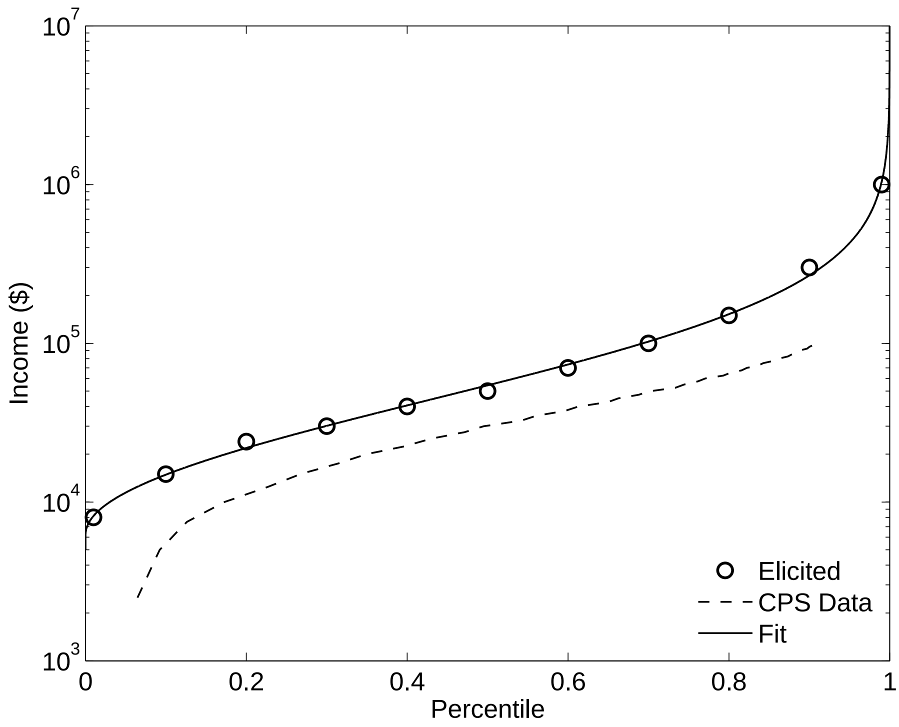


Figure 3.3. Aggregate (median) estimated incomes for each percentile along with offset lognormal fit (solid line) and Personal Income distribution data from the Bureau of Labor Statistics and the Census Bureau Current Population Survey (dashed line).

Table 3.1. Elicited median incomes in dollars for all percentiles (k denotes thousands, M millions).

Percentile	1^{st}	10^{th}	20^{th}	30^{th}	40^{th}	50^{th}	60^{th}	70^{th}	80^{th}	90^{th}	99^{th}
Medians 10k+	8k	15k	24k	30k	40k	50k	70k	100k	150k	300k	1M
Medians All participants	7k	15k	24k	30k	40k	50k	70k	100k	150k	300k	1M

Table 3.1 above shows the median estimated income values for each elicited percentile in dollars for all participants in our study as well as for participants with reported income above \$10,000. Note that the median incomes for each percentile are unchanged if all participants are included.

We found that in each income percentile, individuals consistently overestimated the income compared to the CPS personal income distribution (by a factor of approximately 2 between the 20th and 60th percentiles). To evaluate the Gini coefficient of the elicited aggregate income distribution we fitted an offset lognormal distribution to the data. The offset lognormal distribution is given as a regular lognormal distribution, with all values increased by a constant θ so that the PDF is defined on the interval $[\theta, \infty)$. This distribution was fitted to both the median data and the individual-level data (see also Figure 3.5 and Figure 3.6), and the particular parametric form was chosen because, while it is established that lognormal distributions are good approximations to income distributions, the data suggested that many participants' elicited income distributions would be best fit by a distribution whose lowest value was greater than zero. We note that the offset lognormal and standard lognormal distributions form nested models, so for example, an individual that gives an income of 0 for the first percentile would be best fit by a standard lognormal distribution ($\theta = 0$).

The elicited distributions consisted of incomes for eleven pre-determined percentiles, and were thus points on the inverse CDF curve of the income distribution. To estimate the parameters, we fitted our target function (inverse CDF) to the data using non-linear least squares. We found an excellent fit[17] to the median data, with R^2 = 0.9998. We calculated the analytic formula for the Gini coefficient of this distribution starting from the standard formula for the lognormal distribution and adapted it to account for the constant θ (see 3.17.1 Appendix I). The Gini coefficient obtained from this procedure was equal to 0.6163, which is a little greater than the CPS reported Gini of 0.519 for personal income data. Note that this value of 0.6163 was robust to the removal of, for example, the 99th percentile data and or 90th percentile.

[17] Parameter values: $\mu = 10.79, \sigma = 1.32, \theta = 5890$.

We also fitted the personal income CPS data to the offset lognormal distribution, see Figure 3.4.

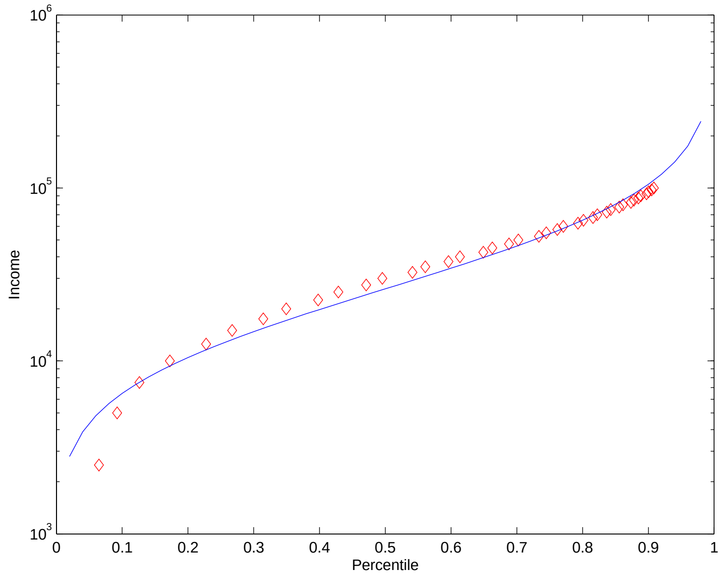


Figure 3.4. Offset lognormal fit (solid line) to Personal Income distribution data from the Bureau of Labor Statistics and the Census Bureau Current Population Survey.

The Gini coefficient for this fit was equal to 0.5578 with an $R^2 = 0.88975$. By removing the first data point from the dataset, we found an improved fit. This was associated with a lower Gini coefficient of 0.5299, and a much better R^2 of 0.95605. Note that this Gini coefficient is higher than the one quoted by the Bureau of Labor Statistics and Census Bureau which is 0.519 but the difference is only 0.01, confirming that our fitting method is valid.

Note that we can evaluate participants' accuracy both for the Gini coefficient and for income values for the different percentiles. If participants gave exactly half the income value for each percentile, Gini would be exactly the same numerically but perceived incomes would be lower. On the other hand, some elicited incomes could be close to actual values while at the same time the Gini coefficient could be different compared to the actual one.

Figure 3.5 shows the aggregate income distribution elicited from participants, computed as the median income in each percentile, along with the personal income distribution for 2015 taken from the Bureau of Labor Statistics and Census Bureau

Current Population Survey (CPS), the CPS 2015 Household income data and the Survey of Consumer Finance (SCF) 2016 data that contains income data from a mixture of households and individuals. We observe that participants gave very close estimates compared to the CPS household and SCF data at least for the middle percentages, although participants were asked to elicit individual incomes and not household incomes. Since we did not also elicit household income distributions from participants it is theoretically possible, although intuitively implausible, that participants could be perhaps arbitrary to deduct that participants were not able to differentiate between personal and household incomes. Note that participants still overestimated high end incomes even when their estimates were compared to household incomes.

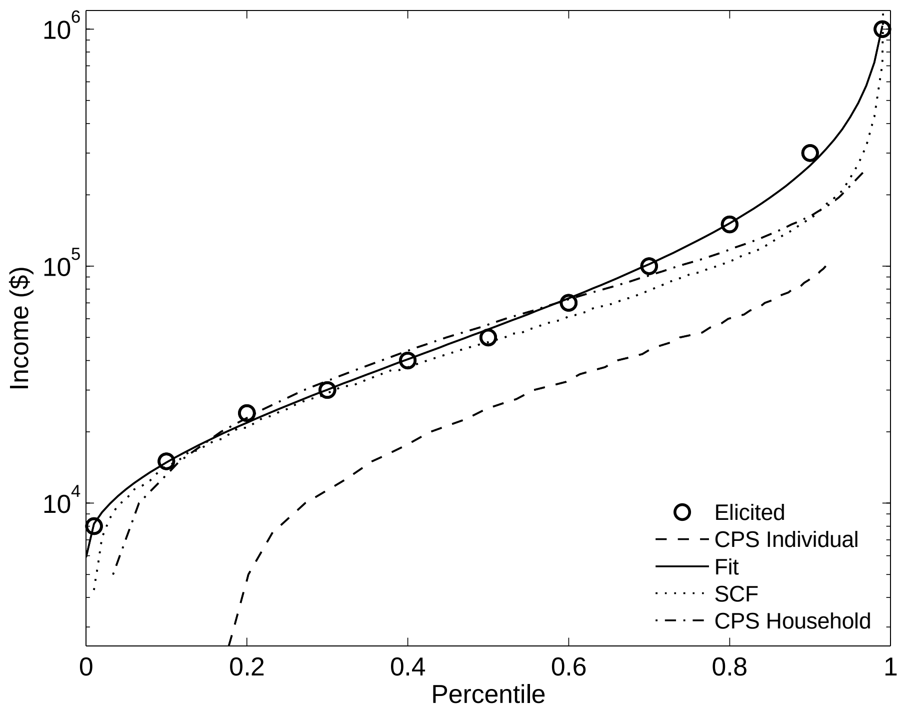


Figure 3.5. Aggregate (median) estimated incomes for each percentile along with offset lognormal fit (solid line), CPS Individual refers to the Current Population Survey data on Personal Incomes (Gini reported by Census 0.519), CPS Household refers to the Current Population Survey data on Household Incomes (Gini reported by Census 0.479), SCF refers to the Survey of Consumer Finance that has income data from a mixture of households and individuals.

3.9.2 Measuring Individual Perceptions of Income Inequality

To measure inequality on an individual level, we estimated the Gini coefficients of the income distributions elicited by participants in the same manner as for the aggregate data. We note that the majority of participants did not give wholly consistent answers (stating that the income of the 40th percentile was lower than the income of the 20th percentile, for example) and because of this not all participants' responses could be well fit, especially for those from whom we elicited a particularly non-monotonic income distribution (as measured by the Kendall tau coefficient). Therefore, to ensure estimates of this Gini coefficient and other quantities derived from this distribution were reliable, we included only those individuals for whom we found an adequate fit. To measure goodness of fit we used the standard R^2 measure. We included only individuals with an R^2 bigger than 0.9 in all subsequent analysis. This resulted in a sample of 551 individuals out of 724 individuals whose income was higher than $10,000[18].

Figure 3.6 shows example fits obtained with this method for participants with a range of distribution shapes and Gini coefficients, demonstrating that this method is flexible enough to capture a variety of distributions.

Figure 3.7 shows the distribution of Gini coefficients for the participants in our study with stated income above $10,000 and $R^2 > 0.9$. The median Gini coefficient was 0.6567 and the mean 0.6669, both of which are slightly higher than the values for the aggregate case. Indeed, the majority (417/551) of individuals' estimated distributions had a Gini coefficient greater than that of the CPS personal income data, showing that individuals consistently overestimated the Gini coefficient.

[18] The analysis was also conducted for different cutoffs of R^2; $R^2 > 0.7$, $R^2 > 0.8$ but such changes did not result in any qualitative differences in these or any of the following results.

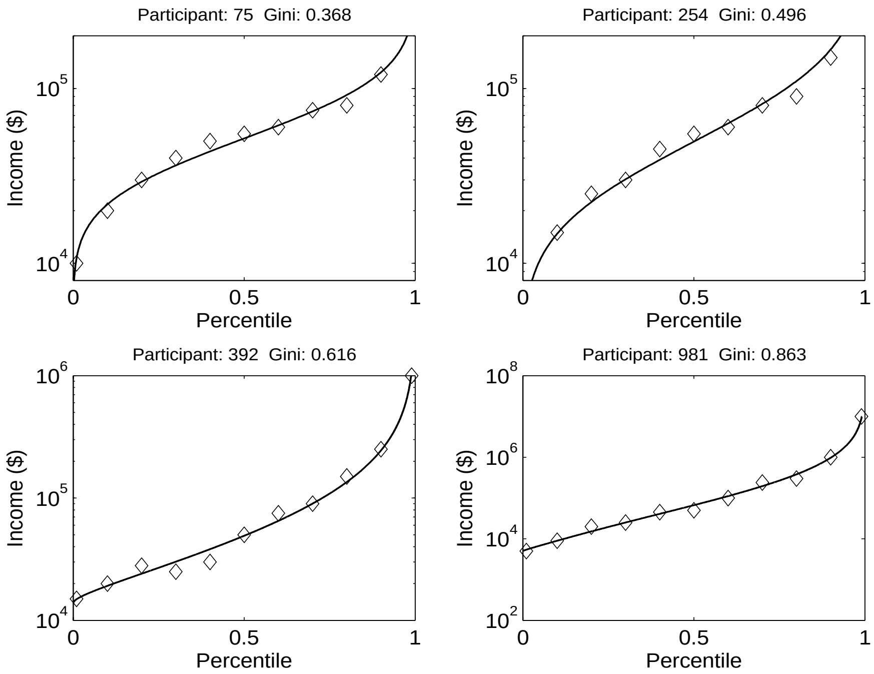


Figure 3.6. Example results of fitting an offset lognormal distribution to elicited Income distributions for four participants with differing levels of perceived Inequality measured by the Gini coefficient.

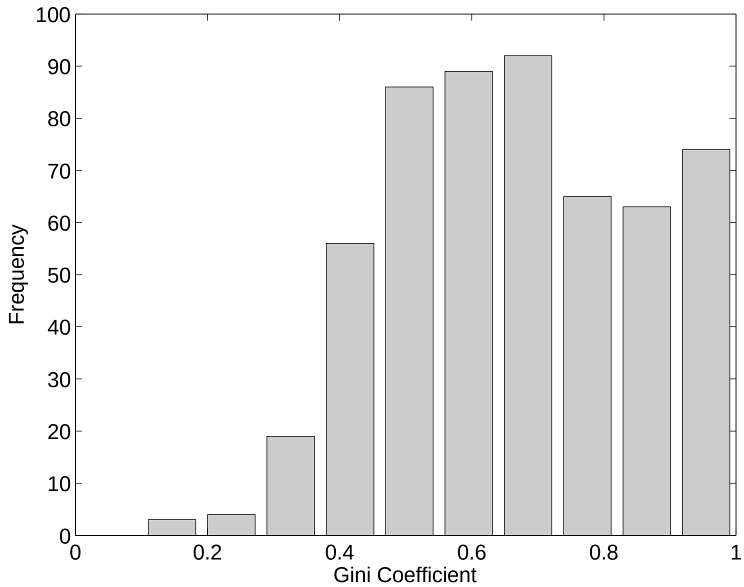


Figure 3.7. Distribution of Gini coefficients for participants with Personal Income greater than \$10,000 and R^2 >0.9. Gini coefficients computed by fitting an offset lognormal distribution to the elicited inverse CDF. Mean Gini coefficient found to be 0.6669, median 0.6567.

3.9.3 Comparing our Measure of Income Inequality

To evaluate the robustness of our method for calculating the Gini coefficient we compared it to participants' estimates of the CEO-worker measure of inequality. However, we do not believe that the CEO-worker measure is likely to be a valid measure of perceived inequality as people may bring to mind, using an availability heuristic, famous examples of CEO's like Bill Gates or Mark Zuckerberg resulting in high overestimations of inequality. In our study individuals' estimates of the income for an average CEO and average worker were elicited. To facilitate easy comparison with the Gini coefficient, we used the following inequality CEO-Worker (CW) index:

$$CW = 1 - \frac{I_{WRK}}{I_{CEO}},$$

equal to one minus the ratio of the income of the worker to the income of the CEO. In this way, perfect equality, where the worker and CEO receive the same income, corresponds to $CW = 0$ and perfect inequality, where the worker earns nothing and the CEO earns everything corresponds to $CW = 1$, as in the case of the Gini Coefficient. We found that the Gini coefficient and this index were positively correlated within an individual ($r_s = 0.260$, $p < .001$). Taking instead the log of the worker's income and the log of the CEO's income, we again found that the Gini coefficient and this index were positively correlated within an individual ($r_s = 0.249$, $p < .001$).

Instead of using the individual Gini coefficient, we also compared the CW ratios of individuals with the ratios of incomes elicited for different percentiles. Three ratios were used to compare with the CW ratios, the 99th to the 1st percentile, the 90th to the 10th percentile and the 80^{th} to the 20th. Correlation coefficients of these ratios and the CW ratio are shown in Table 3.2.[19]

[19] Note that only individuals who estimated that the income of a CEO was equal or greater than that of a worker were used in all the relevant calculations.

Table 3.2. Spearman's correlation coefficients between CEO/Worker ratio and various elicited percentile ratios. Coefficients are all significant at p < .001.

Percentile ratio	99/1	90/10	80/20
Correlation Coefficient	0.274	0.197	0.145

3.9.4 Where Do the Different Perceptions about Percentiles Come From?

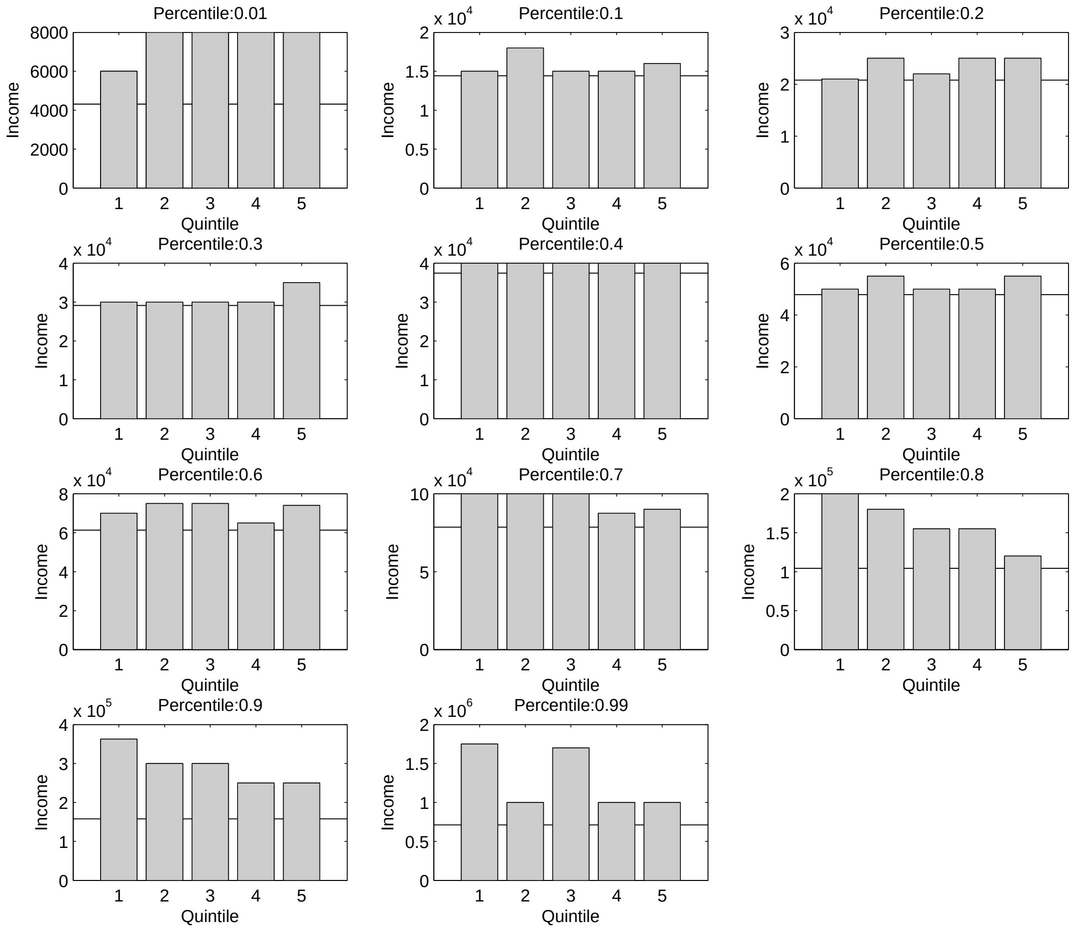


Figure 3.8. Median estimated incomes for different elicited percentiles by equally split Income groups.

There was a degree of heterogeneity in participants' perceptions of inequality. Where do these differences in perceptions come from when it comes to income levels? Each graph in Figure 3.8 corresponds to a different percentile (0.01, 0.2, 0.3 etc. as used in our study). Each of the five groups, or quintiles, corresponds to the individuals with income in the bottom 20%, 20%-40%, 40%-60%, 60%-80% 80%-100% of our sample. This meant that each group had the same number of individuals, based on their personal income. We then took the median elicited income values for each group. The black line is the reference income from the Survey of Consumer Finance (SCF). We used the SCF data as reference here because the CPS personal income and household data give incomes in intervals. The main effect of participant income is evident in the estimates of the high percentiles (0.8, 0.9, 0.99), where we see that group in quintile 1 gave the highest estimates for the income and correspondingly group 4 and 5 gave the lowest estimates. Also, for the first percentile, group 1 gave the lowest estimates compared to the other groups. This implies that lower income individuals should perceive more income inequality which is what we report later (see Table 3.4 and Table 3.5) by using a continuous scale for personal income and avoiding issues of aggregation.

3.9.5 Individual Differences in Perceptions of Income Inequality

Figure 3.9 shows the relationship between individuals' computed Gini coefficients and their reported personal income. We see a negative relationship, with a correlation coefficient of r = 0.110 (p = .009). We find that distributions elicited from higher income individuals displayed lower levels of inequality than those elicited from individuals with lower incomes.

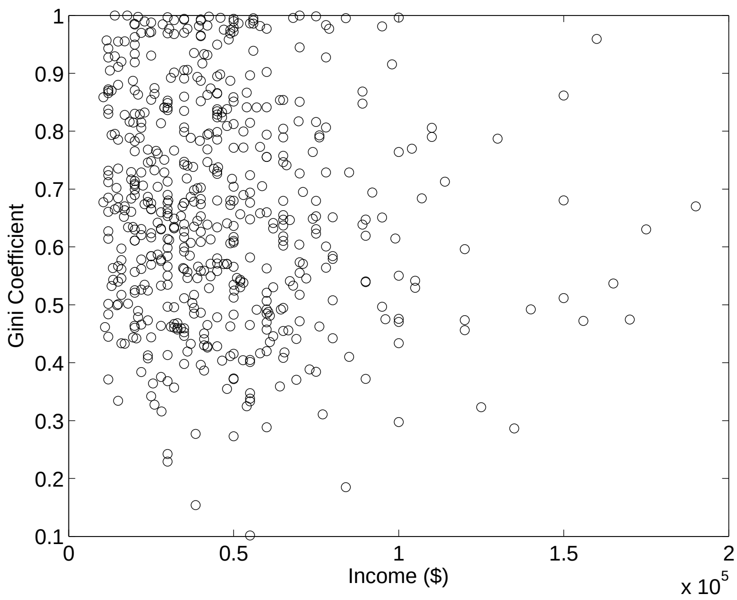


Figure 3.9. Income versus Gini Coefficient.

To understand the effect of individual differences between participants more fully, we performed an analysis incorporating demographic information, age and gender, as well as measures of political ideology. To measure political ideology, we used the Everett scale of conservatism (Everett, 2013), which yields two estimates of conservatism, economic and social, which can be combined to form an overall measure of conservatism. We also used a left-right scale, measuring political ideology from very liberal to very conservative. We note that two individuals did not successfully complete the Everett scale questions and so were excluded from this analysis[20]**.** Table 3.3 shows the correlations between the predictor variables. There were no significant collinearities outside of the measures of political ideology, which we expected to be heavily correlated.

[20] This accounts for the slight difference between the income/Gini correlation coefficient in Table 3.3 and that reported in the text.

Table 3.3. Pearson's (Spearman's) correlation matrix between variables for the regression reported in Table 3.4.

	Age	Gender	Social	Economic	Overall	Left-Right Ideology	Gini
Income	0.032	-0.099	0.116	0.066	0.109	0.091	-0.112
	(0.097)	(-0.122)	(0.128)	(0.092)	(0.123)	(0.124)	(-0.112)
Age		0.089	0.291	0.102	0.248	0.165	0.055
		(0.070)	(0.278)	(0.102)	(0.241)	(0.136)	(0.066)
Gender			0.111	-0.046	0.061	-0.087	0.038
			(0.107)	(-0.063)	(0.051)	(-0.096)	(0.039)
Social				0.573	0.944	0.674	-0.106
				(0.567)	(0.944)	(0.662)	(-0.088)
Economic					0.812	0.701	-0.115
					(0.799)	(0.698)	(-0.120)
Overall						0.763	-0.122
						(0.752)	(-0.111)
Left-Right							-0.126
Ideology							(-0.110)

The results of an OLS regression of individual Gini coefficients against the demographics, income and political ideology measures are shown in Table 3.4 with the results for an equivalent quantile regression shown in Table 3.5. Apart from gender, the estimated coefficient for each predictor was found to be significant. The results of Figure 3.9 are reproduced — we once again see that higher income is associated with a lower measured Gini coefficient. Furthermore, we find that higher levels of conservatism are associated with a lower measured Gini coefficient. The magnitude of the OLS regression coefficient for the overall conservatism measure is -0.001288, meaning that, as measured by the Everett scale, an extreme liberal (overall conservatism = 0) and an extreme conservative (overall conservatism = 100) differ in their perceived Gini coefficients by approximately 0.13. As an example, this difference is approximately equal to the difference in the Gini coefficient of household income between the USA (45.0) and Canada (32.1) for 2007 (values taken from the CIA World Factbook[21]).

*Table 3.4. OLS regression results for Gini coefficient (*** denotes p <.05, standard errors in parentheses).*

	Model 1	Model 2	Model 3	Model 4	Model 5
Income	-6.565e-07***	-6.507e-07***	-7.042e-07***	-6.520e-07***	-6.829e-07***
	(2.863e-07)	(2.865e-07)	(2.850e-07)	(2.857e-07)	(2.850e-07)
Age	1.388e-03***	1.457e-03***	1.093e-03*	1.420e-03***	1.251e-03***
	(7.029e-04)	(7.014e-04)	(6.771e-04)	(6.925e-04)	(6.825e-04)
Gender	1.030e-03	1.315e-02	6.276e-03	1.103e-02	3.806e-03
	(1.648e-02)	(1.635e-02)	(1.630e-02)	(1.626e-02)	(1.633e-02)
Social Conservatism	-6.528e-04*	-9.764e04***			
	(4.253e-04)	(3.477e-04)			
Economic Conservatism	-6.708e-04		-1.123e-03***		
	(5.102e-04)		(4.173e-04)		
Overall Conservatism				-1.288e-03***	
				(4.164e-04)	
Left-Right Ideology					-1.464e-02***
					(4.917e-03)
R^2 Multiple	0.0336	0.03053	0.0294	0.03346	0.03225
R^2 Adjusted	0.02472	0.02341	0.02228	0.02637	0.02515
DoF	545	546	546	546	546

[21] https://www.cia.gov/library/publications/the-world-factbook/fields/2172.html.

*Table 3.5. Quantile regression results for Gini coefficient (*** denotes p <0.05, standard errors in parentheses).*

	Model 1	Model 2	Model 3	Model 4	Model 5
Income	-7.974e-07***	-7.927e-07***	-7.763e-07**	-7.962e-07***	-8.669e-07***
	(3.983e-07)	(3.983e-07)	(3.987e-07)	(3.96126e-07)	(4.008e-07)
Age	0.00120	0.00153*	0.00115	0.00129	0.00160*
	(0.00099)	(0.00203)	(0.00098)	(0.00101)	(0.00102)
Gender	0.01297	0.01991	0.01154	0.01751	0.00489
	(0.02639)	(0.02596)	(0.02576)	(0.49561)	(0.02595)
Social Conservatism	-0.00007	-0.00100**			
	(0.00068)	(0.00057)			
Economic Conservatism	-0.00200*		-0.00201***		
	(0.00082)		(0.00068)		
Overall Conservatism				-0.00146***	
				(0.00068)	
Left-Right Ideology					-0.01658***
					(0.00811)
Koenker and Machado R^2 [22]	0.0237269	0.01634415	0.02371589	0.01968813	0.01834235
DoF	545	546	546	546	546

[22] Goodness of fit measure defined in Koenker and Machado (1999).

While conservative individuals perceive there to be less inequality, these results did not extend to political allegiance. A one-way ANOVA was run, examining the difference in Gini coefficients between participants of different political allegiance (Republican, Democrat, Independent and No Preference). No significant differences were found (F (3, 547) = 1.19, p = 0.314) between their perceived Gini coefficients.

3.9.6 Measuring Inferred, Subjective and Objective Ranks of Income

Our study elicited both an individual's perceived income distribution and their individual income. With our fitted distributions we were therefore able to compute an inferred income rank for each individual by computing their rank position within the fitted CDF. To calculate the rank, let x_i be the income of individual i, and F the fitted offset lognormal CDF; then the rank of individual i is given by the formula

$$r_i = F(x_i).$$

We will denote this measure inferred rank. We also elicited a rank measure explicitly for each participant, asking where they felt their income ranked within the countrywide distribution of personal incomes. We term this measure the subjective rank. The subjective rank is given in 12 steps. If we suppose that the first step corresponds to a rank of 0.01 and the final step corresponds to a rank of 0.99, then we have 10 steps to fit the middle. These each were separated by a spacing of 0.98/11 = 0.0890909 and we thereby obtained a set of 12 ranks. Finally, using the CPS personal income data we computed an objective rank for each individual, quantifying their actual rank position within the income distribution of the USA by matching the reported income of each individual to the corresponding percentile interval on the curve in Figure 3.3. Because the CPS personal income data do not record detailed information about income above $100,000, we could not compute detailed objective rank information for individuals with income above this value, and we set their rank to be equal to the rank that corresponded to an income of $100,000. 33 participants in our study reported an income above $100,000.

Histograms of the three rank measures are shown in Figure 3.10. We see that while inferred and subjective ranks have similar shapes, the histogram of objective rank is different. A one-way ANOVA showed that the means of these measures were indeed significantly different (F (2, 1650) = 233.27, p < .01). Post-hoc t-tests revealed that this difference came from the fact that the mean of the objective rank (0.6163) was significantly different ($p < .001$) from both the subjective rank (0.3942) and the inferred rank (0.4030). The means of subjective and inferred rank were not found to be significantly different.

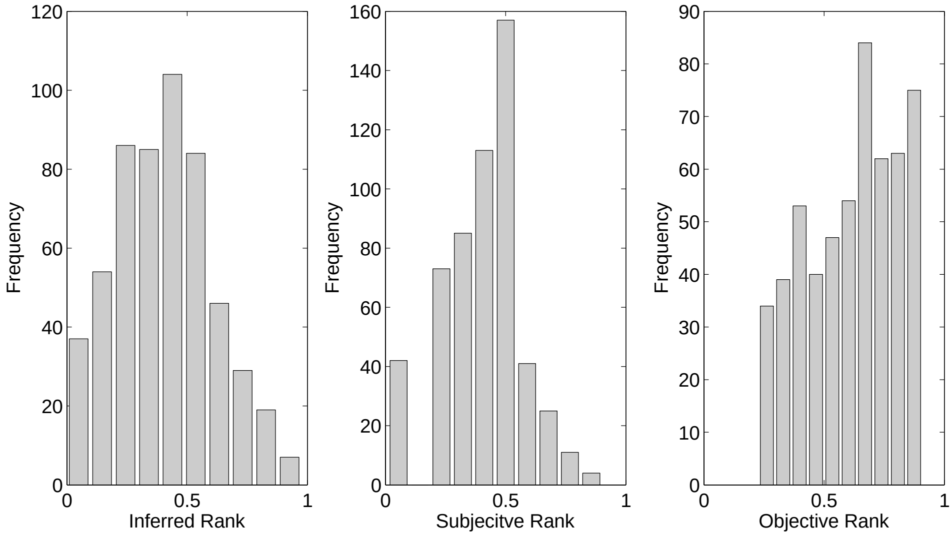


Figure 3.10. Histograms of the three Rank measures.

3.9.7 Evaluating our Measure of Income Inequality

One way to evaluate our methodology of measuring perceptions of inequality is to look at the relationship between subjective rank and inferred rank. Strong correlations would indicate that our methodology captured in fact an individual's internally stored beliefs about the income distribution.

Figure 3.11 illustrates the relationship between subjective and inferred rank, where we observe a strong positive linear relationship, further confirming the strength of the inferred rank measure. Subjective and inferred rank were also found to be highly correlated with personal income. These indicate that our methodology and the associated inferred rank are robust and valid measures.

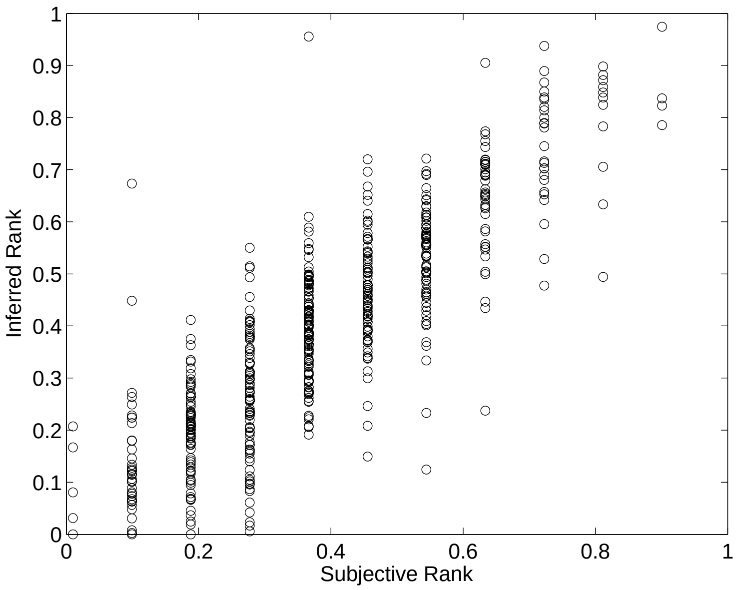


Figure 3.11. Subjective versus Inferred Rank.

Table 3.6 shows the correlation matrix for the three measures of rank, as well as personal income. All measures were found to be highly correlated with each other (all correlations significant at $p < .01$). We note that because objective rank and personal income give the same ranking of individuals (if individual A has greater income than individual B, then individual A's objective rank is also greater than individual B), the Spearman's correlation coefficient between subjective rank and personal income is the same as that between subjective rank and objective rank (the same holds for the inferred rank measure).

Table 3.6. Correlation matrix between Inferred Rank, Subjective Rank, Objective Rank and Personal Income. All correlations were significant at p < .01. Pearson's correlation coefficients with Spearman's in parentheses.

	Inferred Rank	Personal Income	Objective Rank
Subjective Rank	0.8489	0.7498	0.8006
	(0.8488)	(0.8221)	(0.8226)
Inferred Rank		0.7897	0.8388
		(0.8533)	(0.8536)
Personal Income			0.8658
			(0.9991)

3.9.8 Ranks, Income, Income Inequality and Analysis of Well-Being

We turn to the analysis of individuals' subjective well-being data. What best predicts different aspects of well-being? Is it income, subjective rank, inferred rank or perceived Gini? In the present section, we examined which of the rank measures (subjective, inferred, objective), income or income inequality (measured by an individual's perceived Gini coefficient we calculated previously) is better at predicting various measures of subjective well-being. We also considered the logarithm of income as a potential predictor because people's utility might increase as a function of income but not do so linearly. We used six measures of well-being: overall life satisfaction, eudemonic well-being, measures of individual's happiness and anxiety on the day prior to the study, as well as two ladder estimates (one representing one's life from best to worst, the second one's life in about five years from now). Figure 3.12 shows the histograms for all the subjective well-being measures (with only anxiety being heavily negatively skewed) and Table 3.7 shows the correlations of the subjective well-being measures and some of the predictor variables.

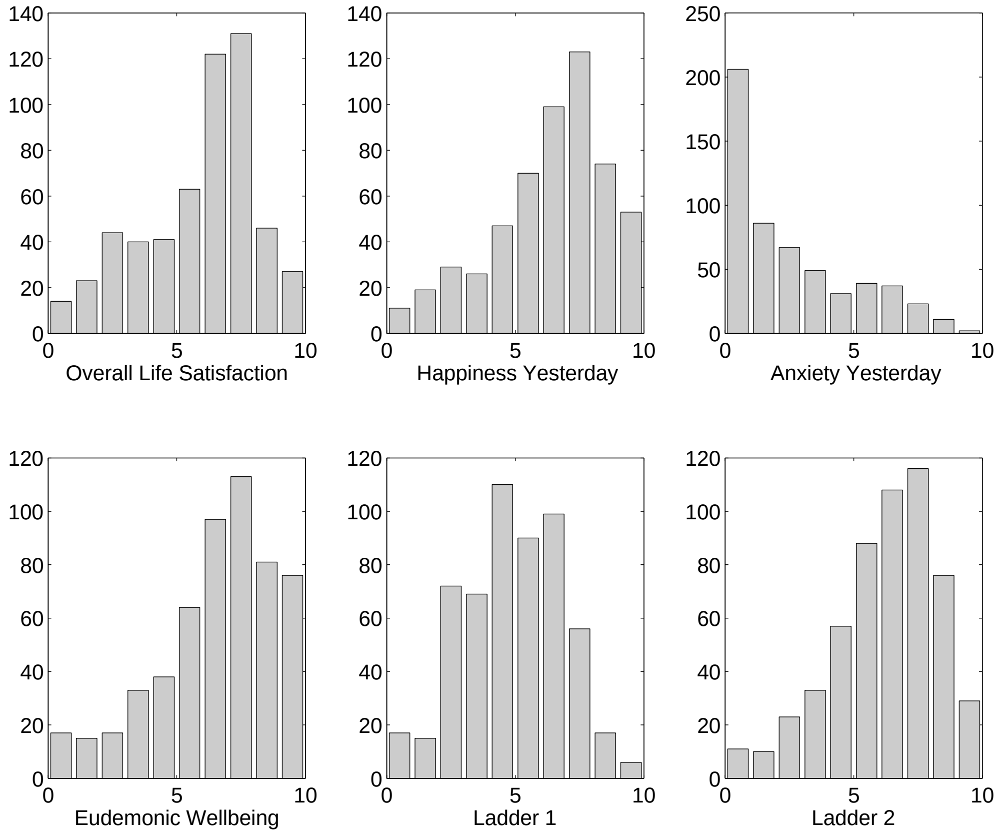


Figure 3.12. Histograms for all Subjective Well-being measures.

Table 3.7. Pearson’s (Spearman’s) correlations between Well-being measures (correlations above 0.5 in bold).

	Happiness Yesterday	Anxiety Yesterday	Eudemonic Well-being	Ladder 1	Ladder 2	Personal Income	Gini	Age	Gender
Overall Life Satisfaction	**0.792**	-0.497	**0.774**	**0.697**	**0.556**	0.247	-0.066	0.074	0.049
	(0.773)	(-0.494)	**(0.751)**	**(0.667)**	**(0.511)**	(0.301)	(-0.075)	(0.086)	(0.053)
Happiness Yesterday		**-0.583**	**0.703**	**0.552**	0.479	0.184	-0.039	0.068	0.041
		(-0.570)	**(0.679)**	**(0.528)**	(0.441)	(0.225)	(-0.042)	(0.058)	(0.044)
Anxiety Yesterday			-0.452	-0.320	-0.223	-0.054	0.004	-0.122	-0.018
			(-0.458)	(-0.288)	(-0.206)	(-0.101)	(-0.003)	(-0.107)	(-0.018)
Eudemonic Well-being				**0.602**	**0.610**	0.210	-0.048	0.110	0.059
				(0.564)	**(0.540)**	(0.216)	(-0.046)	(0.117)	(0.049)
Ladder 1					**0.807**	0.409	-0.091	0.092	0.034
					(0.807)	(0.439)	(-0.106)	(0.102)	(0.029)
Ladder 2						0.333	-0.161	-0.038	0.048
						(0.334)	(-0.1812)	(-0.010)	(0.056)
Personal Income							-0.112	0.032	-0.099
							(-0.112)	(0.097)	(-0.122)
Gini								0.055	0.0381
								(0.066)	(0.039)
Age									0.089
									(0.070)

To identify which of subjective rank, inferred rank, objective rank, income and perceived Gini best predicted each of these well-being measures, we ran a regression for each well-being measure, rank/income/perceived Gini pair, with each regression controlling for both age and gender. Note that Delta Rank is the difference between subjective and objective rank[23]. The rank/income/perceived Gini with the associated highest R^2 was then considered as the best predictor for each of the subjective well-being measures.

The results of this procedure are shown in Table 3.8. Although income had a significant effect on all the measures besides anxiety, log of income seems to be a better predictor than income. Moreover, all rank measures were strong significant predictors of all the subjective well-being measures except anxiety. All five rank/income measures had the same relationship (in terms of direction) with each of the well-being measures. Gini on the other hand had the opposite relationship. We observe that perceived Gini had negative marginal effects on the evaluative measures of well-being (overall life satisfaction, ladder 1) and a significant effect on ladder 2; increasing Gini by 0.1 led to a 0.08 decrease in overall life satisfaction, 0.1 decrease in scores on ladder 1 and a 0.2 decrease in ladder 2. Nevertheless, Gini had no effect on any of the affect measures of well-being.

We clearly see that subjective rank outperformed all other variables in predicting the subjective measures of well-being reported by our participants (based on the R^2s). All well-being measures, except for anxiety, had a positive relationship with subjective rank. Increasing subjective rank by 0.1 increases overall life satisfaction by 0.5, eudemonic well-being by 0.4, Happiness by 0.3, scores on ladder 1 by 0.6 and ladder 2 by 0.5. The relationship between subjective rank and anxiety was in the other direction; increasing subjective rank by 0.1 decreased reported levels of anxiety by 0.2.

[23] If the Delta Rank difference is positive, subjective rank is bigger than objective rank, an individual thinks she is more highly ranked than she is. If it is negative, then the individual thinks that her rank is lower than it actually is. As shown in Table U2 in Appendix II, higher income individuals thought that their rank was lower than it actually is. Also, there was a marginal effect of gender with females showing a stronger tendency to report a rank lower than their actual one.

Table 3.8. OLS Regression results for measures of Well-being against Income, Rank measures, perceived Gini etc. Multiple R^2 reported, DoF=547 for all regressions. Independent variables inserted separately in the regressions. Controlling for Age and Gender in each regression.

Well-being measure		Income	Log(1+Income)	Subjective Rank	Inferred Rank	Objective Rank	Delta Rank	Gini
Overall Life Satisfaction	Coefficient	1.996625e-05	1.164565	4.644026	3.168968	3.662401	0.96421	-0.86625
	p value	< .001	< .001	< .001	< .001	< .001	0.233	**0.0879**
	R^2	0.06942968	0.09826081	**0.14325632**	0.08669727	0.1048	0.01006	0.01275
Eudemonic Well-being	Coefficient	1.764647e-05	0.9823169	3.789083	2.598100	3.023970	0.693194	-0.68097
	p value	< .001	< .001	< .001	< .001	< .001	0.4062	0.1940
	R^2	0.05964740	0.07474522	**0.09878612**	0.06417315	0.07635	0.01579	0.01758
Happy Yesterday	Coefficient	1.482658e-05	0.8342541	3.235752	2.024472	2.586600	0.580915	-0.53559
	p value	< .001	< .001	< .001	< .001	< .001	0.471	0.290
	R^2	0.04042492	0.05293951	**0.07240561**	0.03858074	0.05489	0.006962	0.008049
Anxiety Yesterday	Coefficient	-4.72668e-06	-0.3783597	-1.718013	-1.212304	-1.34554	-0.38106	0.042878
	p value	0.2244	0.0433	0.0048	0.0262	0.0179	0.6791	0.9410
	R^2	0.01749782	0.02217425	**0.02906132**	0.02371321	0.02489	0.01515	0.01485
Ladder 1	Coefficient	2.845537e-05	1.474790	5.824579	4.398112	4.500019	1.453266	-1.02851
	p value	< .001	< .001	< .001	< .001	< .001	0.0387	**0.0198**
	R^2	0.1751950	0.2012015	**0.2906886**	0.2104133	0.2029	0.0173	0.01937
Ladder 2	Coefficient	2.466047e-05	1.259699	5.031715	3.773918	3.815927	1.44228	-1.74634
	pvalue	< .001	< .001	< .001	< .001	< .001	0.0485	**0.00013**
	R^2	0.1197713	0.1341049	**0.1991724**	0.1416077	0.1333	0.01119	0.03035

Since subjective rank[24] was found to be the best predictor of all well-being measures, we used it over the other measures in order to run regressions with each of the well-being measures as dependent variables and subjective rank, Gini, age and gender as independent variables. We found no significant effects for the Gini while controlling for subjective rank on any of the well-being measures except for a very marginal effect on the ladder 2 measure (how individuals felt their life will be in several years' time), the results for which are summarised in Table 3.9.

Table 3.9. OLS regression results for the Ladder 2 measure, p values in parentheses.

	Coefficient	Standard Error
Subjective Rank	4.857	0.442
	(<.01)	
Gini	-0.894	0.419
	(0.033)	
Age	-0.0100	0.007
	(0.129)	
Gender	0.537	0.160
	(<.01)	
R^2 Multiple	0.2058	
DoF	546	

[24] We also investigated a nonlinear relationship between subjective rank and overall life satisfaction, testing for effects of last place aversion (not liking being ranked very high or very low compared to other ranks, with no such effects found, see Figure U1 and Table U1 in Appendix II).

3.10 Study 2-Income Inequality Re-test

The goal for this study was to see how or if our Gini measure of perceived inequality changed over time as well as to check for changes in other individual measurements between the original (Study 1) and the re-test study (Study 2).

3.10.1 Methodology

3.10.1.1 Design

In this study we followed the same design as in Study 1. We contacted the participants who took part in Study 1, 10-12 months after they had completed the first study.

3.10.1.2 Participants

The experiment was conducted again using the Amazon Mechanical Turk platform. Our sample initially consisted of the 1003 participants who took part in our first study, with 481 out of 1003 willing to take part in the second study. The experiment lasted on average 30 minutes and participants were paid $1.50 for completion.

3.10.1.3 Procedure

The procedure was the same as in Study 1. As a reminder the main task involved participants seeing a series of these pictures for the different percentiles 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 99. The order of the questions was randomised for each participant. The elicited data were therefore in the form of an inverse cumulative distribution function for each participant. From these data we were then able to compute each participant's perception of inequality in the form of the Gini coefficient as well as calculating the rank position of their own personal annual income in their perceived income distribution (inferred rank). We also elicited participants' subjective rank of income. Prior to the main task we also elicited the same measures of subjective well-being, ideology, demographics as in Study 1.

3.11 Results

3.11.1 Measuring Aggregate Perceptions Over Time

Aggregate information for the re-test study, is shown below in Table 3.10 and Figure 3.13. No exclusions were applied here based on income or R^2. There was an aggregate insignificant increase for the first, 50th, 60th, 80th and 90th percentiles compared to Study 1.

Table 3.10. Elicited median incomes in dollars for all percentiles (k denotes thousands, M millions) between Study 1 and Study 2.

Percentile	1st	10th	20th	30th	40th	50th	60th	70th	80th	90th	99th
Medians Re-test Study 2 (481 participants)	7.5k	15k	25k	30k	40k	54k	75k	100k	175k	350k	1M
Medians Original Study 1 (1003 participants)	7k	15k	24k	30k	40k	50k	70k	100k	150k	300k	1M

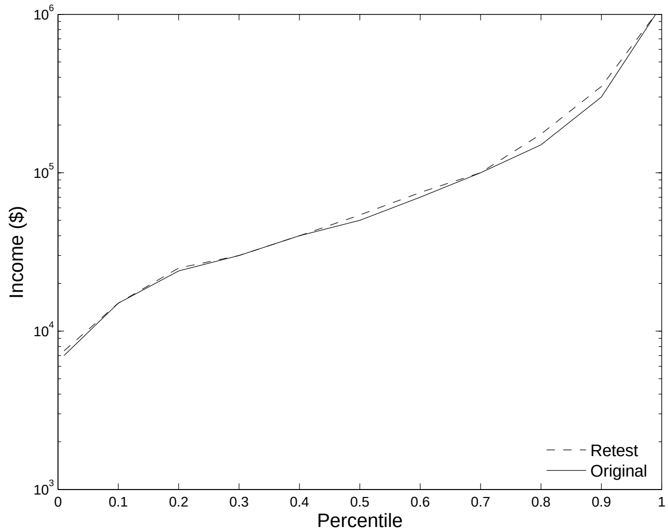


Figure 3.13. Comparison between Study 1 aggregate (median) elicited incomes for each percentile (solid line) and Study 2 aggregate (median) elicited incomes for each percentile (dashed line).

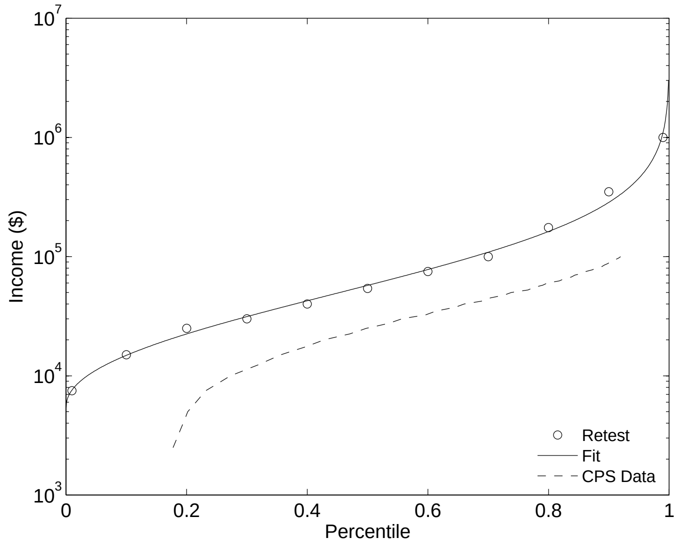


Figure 3.14. Aggregate (median) elicited incomes for each percentile along with offset lognormal fit (solid line) and personal income distribution data from Bureau of Labor Statistics and the Census Bureau Current Population Survey (dashed line).

Overall these values were very close, suggesting that the re-test was consistent with the original study, further validating our methodology. We plot Figure 3.14 and fitted an offset lognormal distribution, finding an R^2 of 0.9956 and a Gini of 0.6458, which is comparable to the original Gini of 0.6163.

3.11.2 Changes Over Time - Looking for Changes in Measures between Study 1 and Study 2

We also looked at how individual perceptions have changed during the months between the experiments. This was a comparison analysis, the goal of which was to compare measurements between experiments. The only purpose was to see if the measurements were consistent. Participants included had income >$10,000 and $R^2 >$ 0.9 in both studies resulting in 285 out of 481 who completed the re-test study.

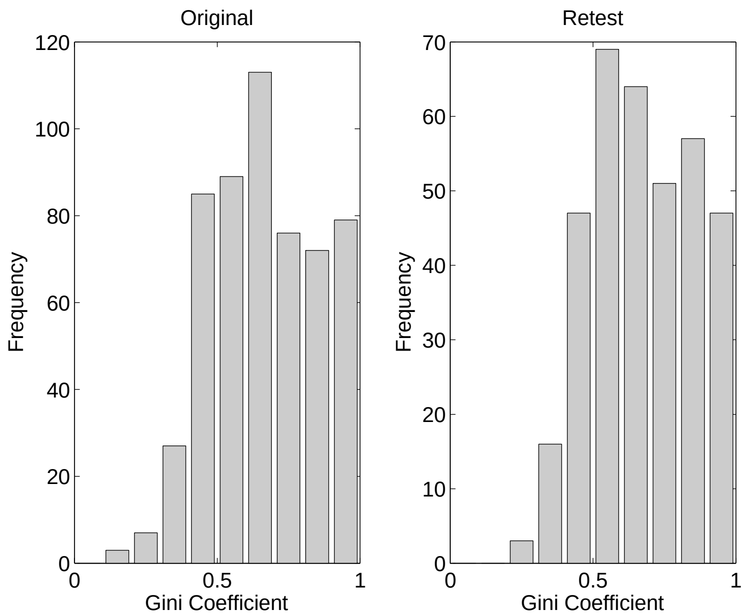


Figure 3.15. Histograms of Ginis; comparison between Original study and Re-test (Income>$10,000, R^2>0.9 in both).

We can see from Figure 3.15 that the shapes of the histograms are very similar between the two studies. Indeed, we found a correlation r = 0.5275, p < .001 between the Gini coefficients across studies and no significant differences among these, t (284) = -0.64255, p = 0.521.

We further looked for changes between other measures we elicited from participants in both studies such as the CW Index, conservatism, overall life satisfaction. We found no significant differences for these measures between the studies. Similarly, our rank measures showed no significant differences between the two studies. Subjective rank in Study 1 correlated significantly with subjective rank in Study 2, r = 0.835, p < .001. Inferred rank's correlation between Study 1 and Study 2 was r = 0.752, p < .001.

The only measure that changed significantly for individuals between the two studies was personal income. Study 2 revealed an aggregate increase in personal income, Wilcoxon signed rank test, V (282) = 7068.5, p < .001, which could be expected given annual raises etc.

3.11.3 Did Changes in Personal Income Influence Changes in Subjective Rank of Income?

Asking our participants questions relating to their personal income and subjective rank less than a year apart, provided us with the ideal means to test if changes in personal income would influence changes in participants' subjective rank of income. To our knowledge this is the first test for this relationship.

We computed a Delta subjective rank = subjective rank of an individual in the re-test study – subjective rank of an individual in the original, and Delta personal income = personal income of an individual in the re-test study – personal income of an individual in the original. We found that a dollar increase in income between the two studies increased Delta subjective rank by 0.00001, see Table 3.11 below. To account for two outliers that showed dramatic changes in income between the two studies we also used a log transformation of incomes to decrease the skewness in the data. This specification showed that increasing log of income by 1% increased Delta subjective rank by 0.0009 (as a reminder subjective rank is measured on a 0-1 scale).

We found no significant effects of changes in income and conservative measures on changes in participants' perceived Ginis (see Table U3 in Appendix III). Finally, we found no effects of changes in income, log of income, subjective rank, inferred rank, objective rank on changes in any of the subjective well-being measures (see Table U4, Appendix III).

*Table 3.11. OLS regression results for the Delta Subjective Rank of Income response measure, *** denotes p values less than .001. Standard errors in parentheses.*

	Model 1	Model 2
(Intercept)	-5.742e-03	-0.0141420
	(2.627e-02)	0.0256820
Delta Income	1.101e-06***	
	(2.750e-07)	
Delta		0.0939305***
Log Income		(0.0167160)
Age	-4.062e-04	-0.0002985
	(4.962e-04)	(0.0004840)
Gender	1.680e-02	0.0178763
	(1.184e-02)	(0.0115460)
R^2 Multiple	0.06163	0.1083
DoF	281	281

3.12 Discussion Study 1 and Study 2

We found that on aggregate individuals consistently overestimated all percentiles of the income distribution, not just the low ones. While some studies found that on aggregate individuals underestimate inequality (Norton & Ariely, 2011; Eriksson & Simpson 2012) the difference between our estimated Gini coefficient (0.6163) and the actual value is 0.097 which is small compared to that found by other methods, and compared to the degree of underestimation found in Norton and Ariely (2011) (although their study was measuring wealth not income inequality, they used the term income inequality). In addition, Page and Goldstein (2016) who elicited household income PMFs from individuals, found an aggregate estimate for the Gini coefficient of 0.34, 0.139 away from the value of 0.479 reported in the 2015 US Census for household incomes (data: Table HINC-01)[25], suggesting our method promotes more accurate responses from people. We note also that the CPS stated Gini coefficient of 0.519 is likely an underestimate due to extremely high-income individuals not being included in the dataset. However, the goal of our methodology was not to measure perceptions of inequality closest to the actual levels but to measure them reliably. Indeed, on an individual level our method was able to capture a variety of income distributions and their associated Ginis.

One way we evaluated our methodology of measuring perceptions of inequality was to look at the relationship between subjective rank and inferred rank. The inferred rank was computed by inserted a participant's personal income in their elicited income distribution and finding its corresponding rank position. Subjective rank was measured by explicitly asking participants to provide their estimated position in the income distribution. These rank measures correlated positively and strongly suggesting that we indeed captured an individual's internally stored beliefs about the income distribution.

Moreover, we found that perceptions of income inequality as measured by the Gini coefficient in Study 1 did not play a role. They did not have significant strong effects on the six subjective well-being measures we used. Furthermore, subjective rank explained more of the variance of the subjective well-being data, compared to income, log of income, inferred rank, objective rank (although they all had significant effects on all six well-being measures). The biggest effects of subjective rank were for the evaluative measures of well-being (overall life satisfaction, ladder 1) and the aspirational measure (ladder 2) compared to the affect measures (eudemonic well-being, happiness, anxiety). That was true for income as well, confirming Kahneman and Deaton's (2010) finding that income has a greater effect on life satisfaction than on emotional aspects of well-being.

In Study 2 we showed that there were no significant aggregate deviations in the Gini coefficients measured within individuals 10-12 months apart, further validating our methodology. Although our Gini coefficient depended on 11 values and an offset lognormal fit making it complicated at least computationally, we found that the

[25] http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc.html

correlation between the Ginis in Study 1 and Study 2 was high. In addition, we found that changes in participants' personal income between the two studies influenced positively and significantly the changes in their subjective estimates of where they rank in the overall income distribution, capturing for the first time subjective rank's dependency on an individual's income. This is important because Brown et al.'s (2018) finding that income has a bigger effect on life satisfaction in more equal societies was attributed at least partially to the relative social rank position that income can confer; if income increases in an equal society then an individual is able to move higher up the ladder of the income distribution compared to an unequal society where incomes are more spread out and so one would need a larger increase in income to increase their social rank.

3.13 Study 3-Wealth Inequality

This study investigated the perceptions of wealth inequality and not income. Very often in the literature these two economic quantities are treated as the same, although they are not. Participants may as well exhibit different perceptions in these two domains, and our methodology is able to capture perceptions for a range of different domains.

3.13.1 Methodology

3.13.1.1 Design

In this study we followed the same design as in Study 1 but converted our questions in order to elicit wealth values for the different percentiles and not income. Everything else remained the same.

3.13.1.2 Participants

The experiment was conducted using the Amazon Mechanical Turk platform. Our sample consisted of 1000 participants subject to a Unites States location filter (M_{age} =37.2, 47.2% female. The experiment lasted on average 30 minutes and participants were paid $1 for completion. Participants who had taken part in Study 1 and Study 2 were not able to take part in this study and we filtered them through their MTurk ID's as well as their IP addresses.

3.13.1.3 Procedure

The procedure was the same as in Study 1. Prior to the elicitation of wealth for the different percentiles, participants were given instructions and a graphical representation of what would follow and what would be asked from them. Moreover, the definition of wealth (taken by the Bureau of Labor Statistics and Census Bureau Survey of income and Program Participation) was provided to participants as follows:

"Wealth, also known as net worth, is defined as the total value of everything someone owns minus any debt that he or she owes. A person's net worth includes his or her bank account savings plus the value of other things such as property, stocks, bonds, art, collections, etc., minus the value of things like loans, mortgages and credit card bills".

The wealth distributions were elicited by asking participants for their estimate of the wealth of 11 different percentiles. An example of such a question is shown below:

"Look at the graphic below. Imagine that it represents the entire population of the USA. All residents are ordered from the poorest (left side of the graphic) to the richest (right side of the graphic) in terms of total wealth. In other words, the leftmost person in the row has the least wealth, whereas the rightmost person has the most wealth.

On the next screen, we will ask you some questions about the wealth of other people."

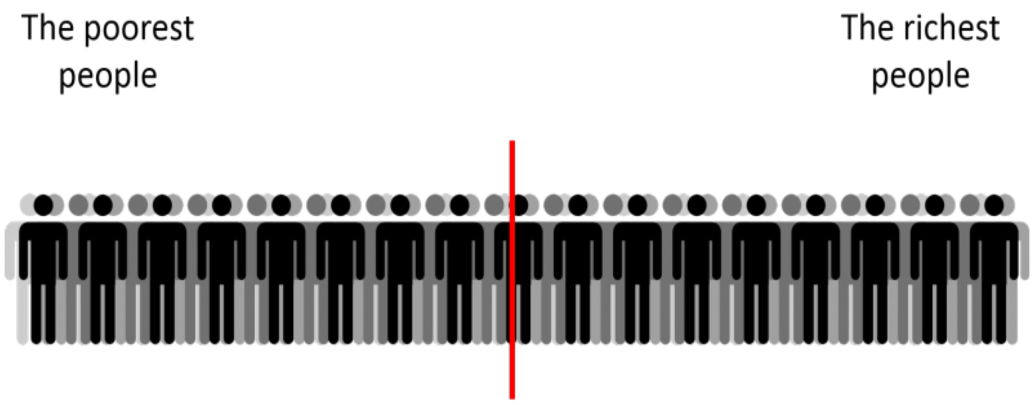


Figure 3.16. Example question shown to participants in Study 2.

"The red line points towards those whose wealth is higher than 50% of the US population, and lower than 50% of the US population.

What do you think the wealth of such a person is?"

Participants saw a series of these pictures for the different percentiles 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 99. The order of the questions was randomised for each participant. The elicited data were therefore in the form of an inverse cumulative distribution function for each participant. From these data we computed each participant's perception of wealth inequality in the form of the Gini coefficient as well as calculating the rank position of their own personal wealth in their perceived wealth distribution. We also elicited subjective measures of the participant's wealth rank by asking them directly to choose into which percentile of the wealth distribution they fell.

As in Study 1, other wealth inequality measures (CW Index for wealth) were also elicited as well as information about subjects' subjective well-being, demographics and political ideology. Again, these were asked at the beginning of the study and prior to the elicitation of distributions. Finally, at the end of this study we asked participants to tell us whom did they vote for in the last US presidential elections. This was a new question not asked in Study 1 and therefore we positioned it at the end of Study 2.

3.14 Results

3.14.1 Measuring Aggregate Perceptions of Wealth Inequality

Table 3.12 below shows the median wealth values for each elicited percentile in dollars for all participants in Study 3. The Census Bureau data reported in the table are 2011 household wealth data updated for inflation (cumulative inflation from 2011 to 2016 was 7.3%[26]). We asked participants for individual wealth not household wealth. The Survey of Income and Program Participation from the Census Bureau does not provide data on individual wealth on their database, only household data. The Census data reported in the table therefore refer to households.

[26] http://www.usinflationcalculator.com

Table 3.12. Elicited median Wealth in dollars for all percentiles (k denotes thousands).

Percentile	1st	10th	20th	30th	40th	50th	60th	70th	80th	90th	99th
Medians All participants	1k	10k	20k	30k	50k	75k	100k	250k	500k	1000k	20000k
Census		-6k		8k		73k		221k		676k	
SCF	-8k	-2k	4k	15k	39k	83k	151k	253k	439k	967k	8000k

The Census data[27] give medians for each quintile. These correspond to percentile values for 0.1, 0.3, 0.5, 0.7, 0.9. The 0.1 percentile is negative for the Census data, so it is not visible on the log plot below (values must be positive for a log plot). We note that people overestimated the wealth of the low percentiles (10th, 30th) and high percentile (90th) compared to the household census data but were accurate for the middle percentiles (50th, 70th).

Since the Census data only contained information on household wealth and solely for quintiles, we also utilised the Survey of Consumer Finance (SCF) because it could provide us with all the percentile values (not just quintiles) and their corresponding wealth values. Moreover, the SCF has collected responses from a mixture of households and single individuals[28]. The primary economic unit in a married couple was one of the two, and for unmarried couples but living together the primary economic unit was the male of the house. For people living alone the primary economic unit was themselves. As the authors warn and advise, in the SCF it is impossible to know which member of the house has what in terms of wealth. The SCF data also helped us with computing an objective rank of personal wealth because the Census household data could only provide us with 4 percentiles. We note that people overestimated the wealth of the low percentiles compared to the SCF data but were accurate for the middle and higher percentiles (except for the 99th percentile). Moreover, they underestimated the wealth of the low percentiles less when compared

[27]https://www.census.gov/people/wealth/data/disttables.html;
https://www.census.gov/data/tables/2011/demo/wealth/wealth-asset-ownership.html.

[28] There were no other available databases that could provide us solely with personal wealth data.

to the Census data (see Figure 3.17). Of course, this is because the SCF data give higher values for wealth at the low percentiles compared to the household Census data.

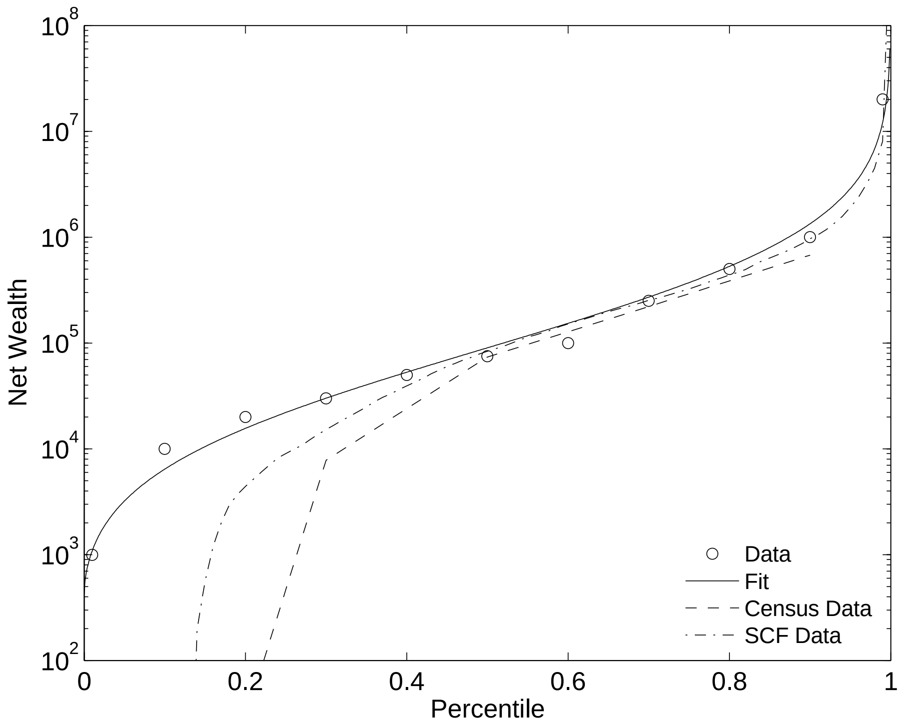


Figure 3.17. Aggregate (median) Wealth for each percentile along with offset lognormal fit (solid line), Census data refers to the Survey of Income and Program Participation data on Household Wealth, SCF refers to the Survey of Consumer Finance that has Wealth data from a mixture of households and individuals.

To evaluate the Gini coefficient of the elicited aggregate wealth distribution we fitted an offset lognormal distribution to the data because again the data suggested that many participants' elicited wealth distributions would be best fit by a distribution whose lowest value was lower (net worth can take negative values) or higher than zero. The elicited distributions consisted of wealth values for eleven pre-determined percentiles, and were thus points on the inverse CDF curve of the wealth distribution. To estimate the parameters, we fitted our target function (inverse CDF) to the data using non-linear least squares. We found an excellent fit, with $R^2 = 0.9878$. The Gini coefficient obtained from this procedure was equal to 0.8636, which is very close to the US Gini for wealth of 0.801 (Davies, Sandström, Shorrocks, & Wolff, 2011).

3.14.2 Measuring Individual Perceptions of Wealth Inequality

In the same manner as for the aggregate data we measured wealth inequality on an individual level. Similar to Study 1, in order to ensure estimates of the perceived Gini coefficient and other quantities derived from this distribution were reliable, we included only those individuals for whom we found an adequate fit. To measure goodness of fit we used the standard R^2 measure. We included only individuals with an R^2 bigger than 0.9 in all subsequent analysis. This resulted in a sample of 592 individuals out of 1000 participants in our study. Figure 3.18 shows example fits for wealth obtained with this method for participants with a range of distribution shapes and Gini coefficients.

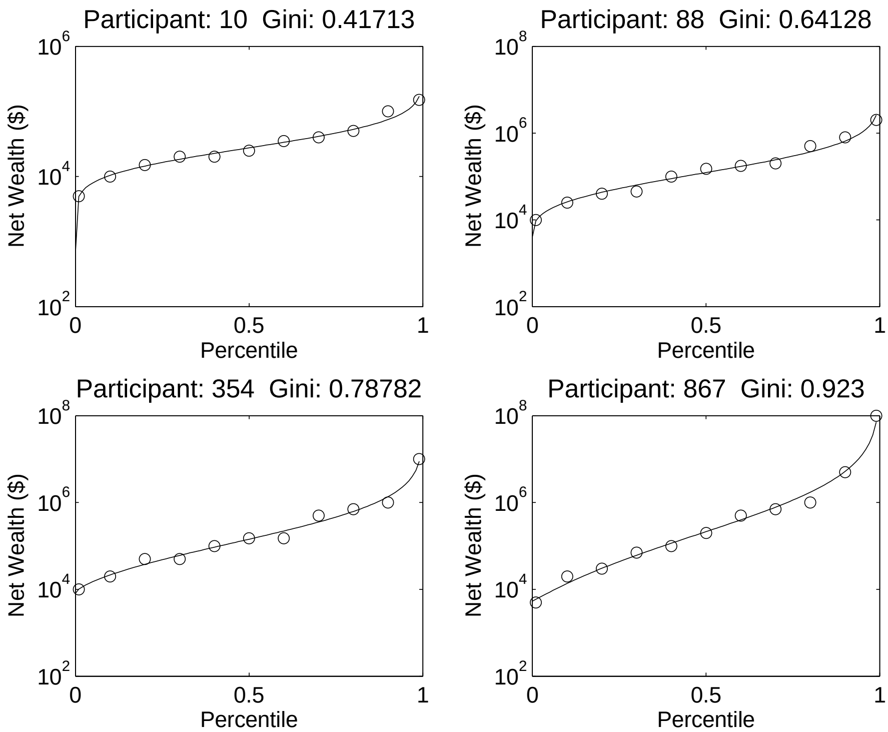


Figure 3.18. Example results of fitting an offset lognormal distribution to elicited Wealth distributions for four participants with differing levels of perceived Inequality measured by the Gini coefficient.

Figure 3.19 shows the histogram of the Gini coefficients. The median Gini coefficient was 0.9360 and the mean was 0.8668 with the histogram being highly skewed to the right. This seems consistent with our aggregate results and the aggregate Gini was equal to 0.8636 (we would not expect these values to be the same as one comes from the aggregate fit, while the other is the mean of all Ginis from the fitted distributions of individuals who had an $R^2 > 0.9$).

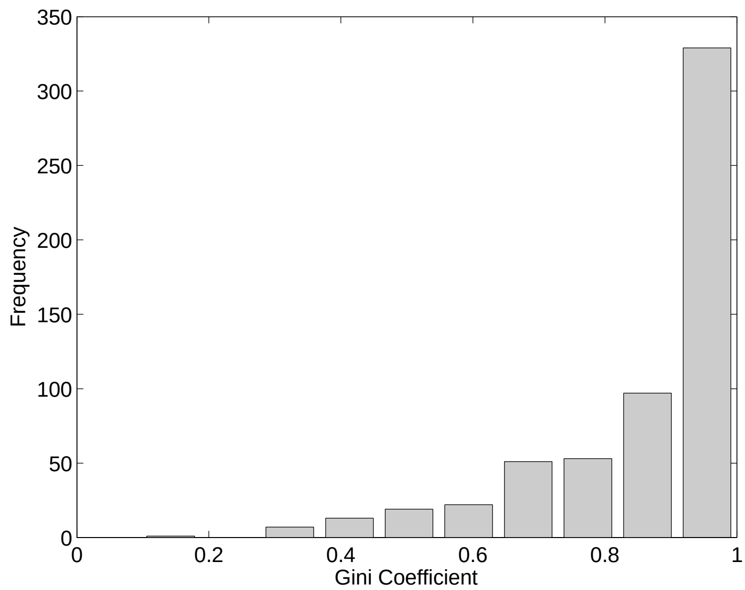


Figure 3.19. Distribution of Gini coefficients for participants with $R^2 > 0.9$. Gini coefficients computed by fitting an offset lognormal distribution to the elicited inverse CDF of Wealth.

3.14.3 Comparing our Measure of Wealth Inequality

To evaluate the robustness of our method for calculating the Gini coefficient for wealth we compared it to the CEO-worker measure of inequality. Participants reported their estimates of the wealth for an average CEO and an average worker. We used again the following inequality CEO-Worker (CW) index:

$$CW = 1 - \frac{I_{WRK}}{I_{CEO}},$$

Many individuals reported the wealth of a worker negative or zero. So, for this section we only considered individuals whose ratios did not become infinite. This resulted in a total of 555 individuals, so 37 were excluded by these criteria.

We found that the Gini coefficient and this index were positively correlated within an individual (r_s = 0.3160, p < .001). We also used the index's logarithmic version which was also positively correlated with Gini (r_s = 0.307, p < .001).

Finally, we looked at the different indices for various ratios of percentiles. Correlation coefficients of these ratios and the CW ratio ($\frac{I_{CEO}}{I_{WRK}}$) are shown in Table 3.13.

Table 3.13. Spearman's correlation coefficients between CEO/Worker ratio and various elicited percentile ratios. Coefficients are all significant at p < .001.

Percentile ratio	99/1	90/10	80/20
Correlation Coefficient	0.3174	0.3068	0.3224

3.14.4 Where Do the Different Perceptions about Percentiles Come From?

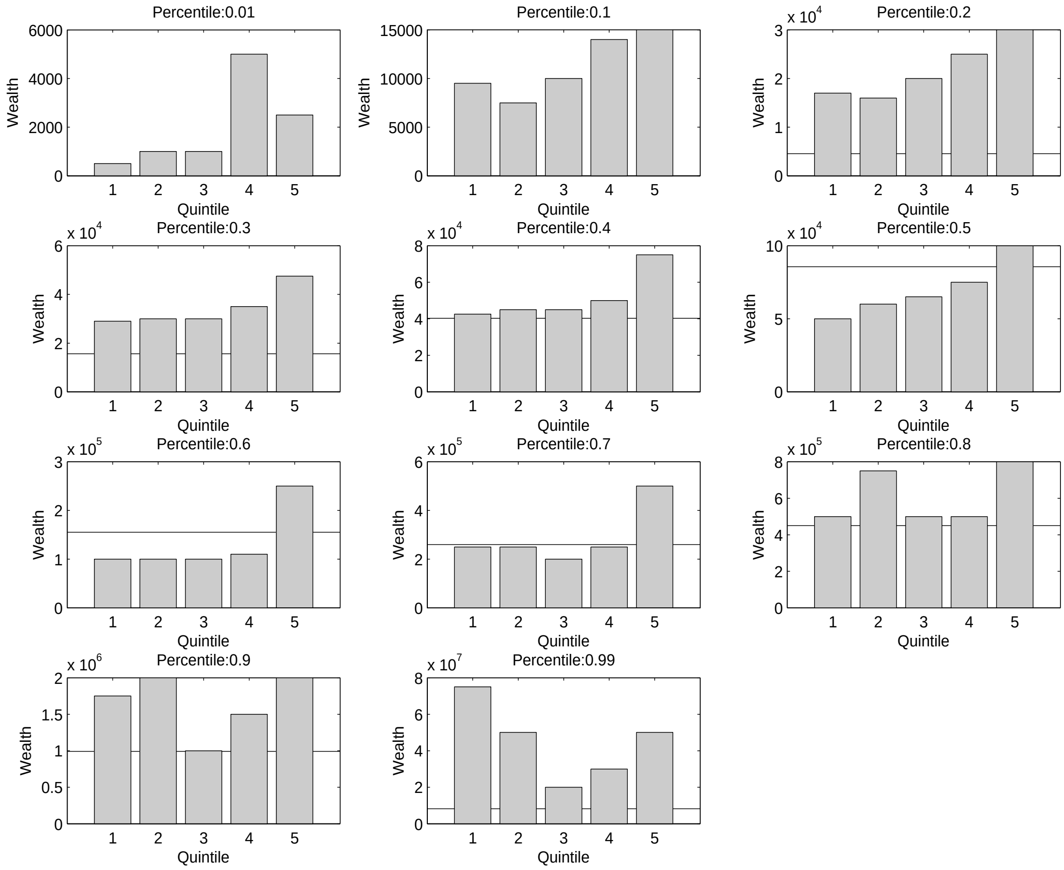


Figure 3.20. Median Wealth for the different elicited percentiles by equally split Wealth groups.

To identify how the perceptions for each percentile were affected by personal wealth levels, we plotted graphs in Figure 3.20 with each corresponding to a different wealth percentile: 0.01, 0.2, 0.3 etc. Each of the five groups (quintiles), corresponded to the individuals with wealth in the bottom 20%, 20%-40%, 40%-60%, 60%-80% 80%-100% of our sample. This meant that each of these 5 groups had the same number of individuals, based on their wealth. We then took the median wealth value for each group. The black line is the reference wealth value from the Survey of Consumer Finance (SCF) data. For the first two percentiles, 0.01 and 0.1, the SCF value is negative and therefore, is not visible in the plot. We observe that quintile 5 is always the highest, except for percentiles 0.8, 0.9 and 0.99. This shows that wealthier individuals overestimated the wealth of the lower percentiles in comparison to the other groups. At the high percentiles, 0.8, 0.9 and 0.99, they were similar to the other groups. This implies that wealthier individuals should perceive lower wealth inequality, which is what we see in Table 3.15 using a continuous scale for personal wealth. We also observe quintile 1 and 2 overestimating the wealth of the 0.99 and 0.90 compared to the 3rd and 4th quintile.

3.14.5 Individual Differences in Perceptions of Wealth Inequality

We investigated individual differences among participants' perceptions of inequality, as measured by the Gini coefficient, in the wealth study also. We performed the same analysis incorporating demographic information; personal wealth, age and gender, as well as measures of political ideology. Table 3.14 shows the correlations between the predictor variables. Again, there were no significant collinearities outside of the measures of political ideology.

Table 3.14. Pearson's (Spearman's) correlation matrix between predictor variables.

	Age	Gender	Social	Economic	Overall	Left-Right Ideology	Wealth Gini
Personal Wealth	0.303	0.040	0.145	0.085	0.137	0.081	-0.018
	(0.340)	(-0.0003)	(0.310)	(0.2168)	(0.304)	(0.229)	(-0.083)
Age		0.140	0.266	0.116	0.234	0.081	0.130
		(0.147)	(0.248)	(0.100)	(0.217)	(0.061)	(0.094)
Gender			0.028	-0.129	-0.032	-0.091	-0.054
			(0.036)	(-0.136)	(-0.027)	(-0.100)	(0.004)
Social				0.582	0.945	0.708	-0.020
				(0.584)	(0.948)	(0.707)	(-0.006)
Economic					0.816	0.658	-0.0036
					(0.806)	(0.666)	(-0.0025)
Overall						0.768	-0.0158
						(0.769)	(-0.0430)
Left-Right Ideology							-0.045
							(-0.068)

The highly skewed residuals (see Appendix IV) in Model 1, 2, 3, 4 using OLS regression diagnostics revealed and suggested that a quantile regression was more appropriate (see also Appendix V for notes on quantile regressions), results of which are shown in Table 3.15.

*Table 3.15. Quantile regression results for Gini coefficient versus Personal Wealth, Age, Gender and Conservatism measures (*** denotes p <.05, standard errors in parentheses).*

	Model 1	Model 2	Model 3	Model 4
Personal Wealth	-8.949710e-08***	-8.718972e-08***	-8.750899e-08***	-9.090003e-08***
	(2.938129e-08)	(948061e-08)	(2.925744e-08)	(2.825611e-08)
Age	2.270117e-03***	2.044558e-03***	2.236549e-03***	1.568714e-03***
	(6.717661e-04)	(6.603550e-04)	(6.685207e-04)	(341120e-04)
Gender	3.965316e-03	-7.875234e-03	-4.126052e-03	1.321324e-03
	(1.459513e-02)	(1.508299e-02)	(1.482956e-02)	(1.442764e-02)
Social Conservatism	-8.023867e-04***			
	(3.304256e-04)			
Economic Conservatism		-3.539783e-04		
		(3.692589e-04)		
Overall Conservatism			-7.017939e-04**	
			(3.795288e-04)	
Left-Right Ideology				-8.845491e-03***
				(4.263349e-03)
Koenker and Machado R^2	0.02123207	0.02123207	0.02200846	0.02123207
DoF	587	587	587	587

The estimated coefficient for each predictor apart from gender was found to be significant (as was observed in Study 1). We see that individuals with higher wealth had a lower measured Gini coefficient. In fact, if personal wealth increases by

$10.000.000, in Model 4, the median Gini would decrease by 0.9, which is a small effect but still significant.

The effect of overall conservatism and left-right ideology is also in the same direction to that of personal wealth, with the effect of overall conservatism being driven mainly by social conservatism. The magnitude of the coefficient for the left-right ideology measure is -0.009, meaning that, as measured by this scale, an extreme liberal (0) and an extreme conservative (10) differ in their perceived Gini coefficients by approximately 0.09.

Because the wealth study was run after the last US elections we also asked participants whom they voted for in the last elections. We then tested for differences between the perceived wealth Ginis among Hilary Clinton and Donald Trump voters. 155 participants stated they voted for Trump, 269 stated they voted for Clinton (summing to 424 out of the 592 sample of participants used in the analysis). We only tested for differences in perceptions of inequality between these two groups as running an Anova between 7 groups (these included other candidates, a did not vote option and a prefer not to say option) of very unequal sizes would yield unreliable results. A Wilcoxon rank sum test revealed no significant differences in the wealth Gini estimates between the two groups (W (423) = 19633, p = 0.3178).

3.14.6 Measuring Inferred, Subjective and Objective Ranks of Wealth

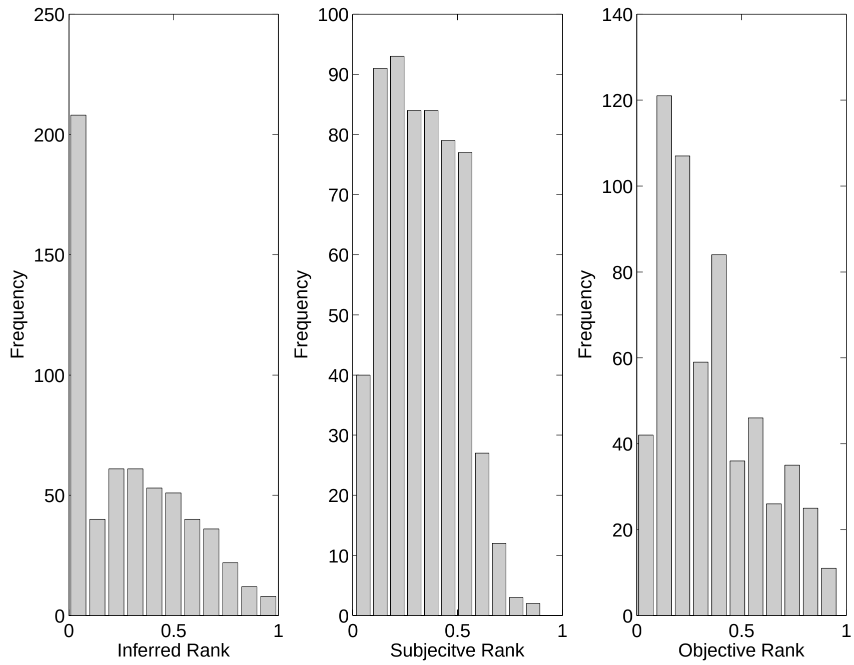


Figure 3.21. Histograms of the three Rank measures.

With our fitted wealth distributions, we computed an inferred wealth rank for each individual by computing their rank position within their fitted CDF. We also elicited a subjective rank measure explicitly for each participant, asking where they felt their wealth ranked within the countrywide distribution of personal wealth. Finally, using the SCF data (because the Census data only provided us with 4 quintiles), we computed an objective rank for each individual, quantifying their actual rank position within the wealth distribution of the SCF data for US individuals and households by matching the reported personal wealth of each individual to the corresponding percentile on the SCF curve in Figure 3.17. Figure 3.21 shows a big difference between inferred rank and the other rank measures, with inferred rank having a big spike at 0. Participants actually placed their own wealth below or equal to their perceived first percentile of wealth (less than half the people in the first bar on the leftmost chart). This means that their inferred rank corresponded to zero, because their wealth would be below the 1st percentile in their fitted distribution. It seems that this phenomenon is associated with the systematic overestimation of low-percentile wealth that we observed in the aggregate statistics.

Nevertheless, when looking at the relationship between subjective rank and inferred rank we found strong correlations, validating our methodology in the case of the wealth study too.

Table 3.16. Pearson's correlation coefficients and Spearman's in parentheses. All correlations were significant at $p < .01$

	Inferred Rank	Personal Wealth	Objective Rank
Subjective Rank	0.7585	0.5058	0.7163
	(0.7411)	(0.7134)	(0.7134)
Inferred Rank		0.5684	
		(0.8708)	
Personal Wealth			0.7004
			(0.9990)

3.14.7 Ranks, Wealth, Wealth Inequality and Analysis of Well-Being

Figure 3.22. Histograms for all Subjective Well-being measures.

Do wealth-related measures predict different aspects of subjective well-being? Is personal wealth, subjective, inferred, objective rank of wealth or perceived wealth Gini better at predicting various measures of subjective well-being? We also considered the logarithm of absolute wealth times the sign function as a potential predictor (log is not defined for negative values which wealth can take; we also added 1 because log of zero is undefined). Participants rated, on a scale from 0 to 10, six measures of well-being: overall life satisfaction, eudemonic well-being, happiness and anxiety on the day prior to the study, as well as ladder 1 and 2 estimates. Figure 3.22 shows the histograms for all the subjective well-being measures (with only anxiety being heavily negatively skewed).

To identify which of the subjective rank, inferred rank, objective rank, personal wealth and perceived wealth Gini best predicted each of these well-being measures, we ran a regression for each well-being measure, rank/wealth/wealth Gini pair, with each regression controlling for both age and gender. The rank/wealth/wealth Gini with the associated highest R^2 was then considered as the best predictor for each of these measures. The results of this procedure are shown in Table 3.17. Personal wealth and the transformation of personal wealth had a positive significant effect on all the measures besides anxiety for which it had a significant negative effect. The transformation of wealth explained more of the variance in all the subjective well-being data compared to wealth. All rank measures were strong significant predictors of all the subjective well-being measures. All five rank/wealth measures had the same relationship, in terms of direction, with each of the well-being measures.

Turning to the effects of perceived wealth inequality as measured by the Gini we observe that Gini had a negative miniscule marginal effect on two of the evaluative measures of well-being (overall life satisfaction, ladder 1), with increases in wealth Gini leading to lower levels of overall life satisfaction and lower scores on ladder 1. Nevertheless, the results were marginal. Moreover, Gini had no effect on any of the affect measures of well-being.

Subjective rank again outperformed all other variables in predicting the subjective measures of well-being reported from our participants (based on the value of the R^2s). All well-being measures, except for anxiety, had a positive relationship with subjective rank, increasing subjective rank by 0.1 increased overall life satisfaction by 0.5, eudemonic well-being by 0.3, Happiness by 0.4, and scores on the ladder 1 by 0.5 and ladder 2 by 0.4. The relationship between subjective rank and anxiety was in the other direction; increasing subjective rank by 0.1 decreased levels of anxiety by 0.26.

Running OLS regressions that included wealth Gini, subjective rank, age and gender as independent variables against overall life satisfaction and ladder 1 made the small effect of wealth Gini disappear but the effect of subjective rank remained. To note here that the correlation between wealth Gini and subjective rank was r = -0.09 and r_s = -0.09.

Table 3.17. OLS Regression results for measures of Well-being against Wealth and Rank measures independently inserted in the regressions. Multiple R^2 reported, DoF=588 for all regressions, controlling for Gender and Age.

Well-being measure		Wealth	Transformed Wealth(sign(wealth) *log(1+|wealth|))	Subjective Rank	Inferred Rank	Objective Rank	Wealth Gini
Overall Life Satisfaction	Coefficient	1.825e-06	0.119772	5.153400	3.133373	4.105101	-0.8514
	p value	<.001	<.001	<.001	<.001	<.001	**0.134**
	R^2	0.04328	0.1067	**0.1903**	0.1324	0.1566	0.009928
Happy Yesterday	Coefficient	1.237e-06	0.083944	3.432816	2.155196	2.937273	-0.6485
	p value	0.00144	<.001	<.001	<.001	<.001	0.2542
	R^2	0.02555	0.05779	**0.08999**	0.06806	0.08528	0.01075
Eudemonic Well-being	Coefficient	1.173e-06	0.084871	3.713935	2.139055	2.899223	-0.17493
	p value	0.00497	<.001	<.001	<.001	<.001	0.77495
	R^2	0.03099	0.06101	**0.09956**	0.06811	0.08194	0.01803
Anxiety Yesterday	Coefficient	-1.176e-06	-0.058629	-2.64538	-1.70935	-2.52371	0.677856
	p value	0.0117	0.00158	<.001	<.001	<.001	0.3204
	R^2	0.02285	0.02885	**0.04571**	0.03814	**0.05144**	0.01387
Ladder 1	Coefficient	2.118e-06	0.114524	5.449680	3.281504	4.127181	-1.1057
	p value	<.001	<.001	<.001	<.001	<.001	**0.024**
	R^2	0.1053	0.1597	**0.3074**	0.2199	0.2375	0.04879
Ladder 2	Coefficient	1.440e-06	0.076049	3.662793	2.069545	2.634737	-0.2669
	p value	<.001	<.001	<.001	<.001	<.001	0.593
	R^2	0.03504	0.05768	**0.1259**	0.07655	0.08551	0.005453

3.15 General Discussion

In this chapter we measured participants perceptions of income and wealth inequality. We measured these perceptions using a new methodology which for a set of 11 percentiles — including the often talked about highest 1% as well as the lowest

1%, participants were asked to estimate the corresponding income/wealth of the US population. These percentiles represent rank positions, which we know from the psychological literature have a heavy influence on decision-making. Our methodology worked well and was validated as in both studies we found that subjective rank was significantly and positively correlated with inferred rank.

Our studies and methodology assisted in differentiating between income and wealth elicited distributions. In Study 1, we found that on aggregate participants overestimated all percentiles, compared to Study 3, where participants overestimated the low and accurately estimated the middle and some of the higher percentiles. The methodology was also able to capture the different individual perceptions of income and wealth, finding that the histograms for income Ginis and wealth Ginis had completely different shapes. Therefore, also on an individual level, participants exhibited different perceptions of income and wealth inequality. Moreover, one of the unifying themes of these two studies was that the same individual differences, personal income/wealth and conservative ideology, affected the perceptions of income and wealth inequality (measured by the Gini coefficient) and these effects were in the same direction in both studies (although these effects were small). It is noteworthy that the effect of personal income/wealth has a different origin in Study 1 and Study 3. High wealth individuals overestimated the wealth of low-wealth individuals while low income individuals overestimated the income of high-income individuals.

Moreover, perceptions of inequality either of income or wealth, measured by an individual's Gini do not seem to matter for well-being. Our methodology may have elicited more reliable and accurate estimates for these perceptions compared to previous studies but individuals seem to not care about them. What they are influenced by, at least for the six subjective well-being measures we gathered responses for, was where participants thought they ranked in the overall income or wealth distribution in their country.

Given that that we found that rank matters it makes sense that a number of studies find no relationship between individual well-being and actual levels of inequality. What is more surprising is that these subjective rank estimates regarded the whole distribution of income or wealth in the US and not that of participants' neighbours or local area. It is possible that individuals sampled from their memory and local surroundings to create these distributions. Nevertheless, we did not have enough data to split participants by state and compare their elicited distributions to the actual distributions in their state.

Future work can look more closely into the perceived Gini and its association with a desired Gini. If perceived Gini did not predict any of the subjective well-being measures strongly in our studies maybe the discrepancy between perceived and desired or objective Gini would be able to shed more light to these findings. A future study can also identify causal relationships in the laboratory by manipulating perceived subjective rank of income and discern if it would create corresponding effects on subjective well-being and preferences for taxation policy.

On a general note, we do not conclude that governments should place less emphasis and efforts on reducing economic inequalities, as its adverse effects on societies have been well documented. We would propose that attention and focus should also be placed on what influences and how to increase one's perception of where they rank in the income/wealth distribution.

3.16 References

Aldrovandi, S., Wood, A. M., Maltby, J., & Brown, G. D. (2015). Students' concern about indebtedness: a rank based social norms account. *Studies in Higher Education*, *40*(7), 1307-1327.

Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics*, *88*, 2009-2042.

Alesina, A., & Giuliano, P. (2011). Preferences for redistribution. In *Handbook of Social Economics* (Vol. 1, pp. 93-131). North-Holland.

Baum, F.C. (Spring 2013). *Quantile regression* EC 823: Applied Econometrics, Boston College. Retrieved from http://fmwww.bc.edu/ECC/S2013/823/EC823.S2013.nn04.slides.pdf

Berg, M., Veenhoven, R. (2010). Income inequality and happiness in 119 nations. In Greve, B. (Ed.), *Social policy and happiness in Europe* (pp. 174–194). Cheltenham, England: Edgar Elgar.

Binelli, C., Loveless, M., & Whitefield, S. (2015). What is social inequality and why does it matter? Evidence from Central and Eastern Europe. *World Development*, *70*, 239-248.

Bjørnskov, C., Dreher, A., & Fischer, J. A. V. (2007). The relation between inequality and happiness-do subjective perceptions matter? *Values and Economy, Proceedings of the 32nd IAREP Conference,* Univerza v Ljubljani, Ljubljana, 471-477.

Bjørnskov, C. (2010), "How Comparable are the Gallup World Poll Life Satisfaction Data?", *Journal of Happiness Studies*, *11*, 41-60.

Bjørnskov, C., Dreher, A., Fischer, J. A., Schnellenbach, J., & Gehring, K. (2013). Inequality and happiness: When perceived social mobility and economic reality do not match. *Journal of Economic Behavior & Organization*, *91*, 75-92.

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471-475.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being?. *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355-389.

Brown, G. D. A., Boyce, C. J., & Wood, A. M. (2018). The wellbeing-income gradient is steeper in more equal countries: Relative rank effects. Unpublished Manuscript.

Brunori, P. (2015). *The perception of inequality of opportunity in Europe* (No. 364). ECINEQ, Society for the Study of Economic Inequality.

Bureau of Labor Statistics and Census Bureau Current Population Survey Data (CPS). (2015). Personal Income Data available online at: http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-01.html. Table PINC-01.

Bureau of Labor Statistics and Census Bureau Current Population Survey Data (CPS). (2015). Household Income Data available online at: http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc.html. Table HINC-01.

Bureau of Labor Statistics and Census Bureau Survey of Income and Program Participation (2011). Household Wealth Data available online at: https://www.census.gov/data/tables/2011/demo/wealth/wealth-asset-ownership.html.

Cansunar, A. (2016). *Calculating with the Unknown: The Effects of (Mis) Perceptions of Inequality Between Income Groups on Tax Preferences.* Unpublished Manuscript. Retrieved from https://sites.duke.edu/preferencesoverredistribution/files/2016/04/Cansunar-inequality-misperceptions.pdf.

Chambers, J. R., Swan, L. K., & Heesacker, M. (2014). Better off than we know distorted perceptions of incomes and income inequality in America. *Psychological Science, 25*(2), 613-618.

Clark, A. E., & Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics, 61*(3), 359-381.

Clark, A. E. (2003). *Inequality aversion and income mobility: A direct test*. Working Paper 2003-11. Delta, Paris.

Clark, A. E., Westergård-Nielsen, N., & Kristensen, N. (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association, 7*(23), 519-527.

Clark, A. E., & Senik, C. (2012). *Income Comparisons in Chinese Villages*. Paris School of Economics. Paris.

Cruces, G., Perez-Truglia, R., & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, *98*, 100-112.

Darvas, Z., & Efstathiou, K. (2016, 9 November). *Income Inequality Boosted Trump Vote.* Retrieved from http://bruegel.org/2016/11/income-inequality-boosted-trump-vote/.

Davidai, S., & Gilovich, T. (2015). Building a more mobile America—One income quintile at a time. *Perspectives on Psychological Science, 10*(1), 60-71.

Davies, J. B., Sandström, S., Shorrocks, A., & Wolff, E. N. (2011). The level and distribution of global household wealth. *The Economic Journal*, *121*(551), 223-254.

Dawtry, R. J., Sutton, R. M., & Sibley, C. G. (2015). Why wealthier people think people are wealthier, and why it matters: From social sampling to attitudes to redistribution. *Psychological Science*, *26*(9), 1389-1400.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth*, 89-125.

Engelhardt, C., & Wagener, A. (2014). *Biased perceptions of income inequality and redistribution* (No. 526). Discussion Paper, Wirtschaftswissenschaftliche Fakultät, Leibniz Universität Hannover.

Engelhardt, C., & Wagener, A. (2016). *What do Germans think and know about income inequality? A survey experiment* (No. 389). ECINEQ, Society for the Study of Economic Inequality.

Eriksson, K., & Simpson, B. (2012). What do Americans know about inequality? It depends on how you ask them. *Judgment and Decision Making*, *7*(6), 741.

Eriksson, K., & Simpson, B. (2013). The available evidence suggests the percent measure should not be used to study inequality: Reply to Norton and Ariely. *Judgment and Decision Making, 8*(3), 395-396.

Everett, J. A. (2013). The 12 item social and economic conservatism scale (SECS). *PloS one*, *8*(12), e82131.

Fahey, A. L., & Smyth, E. (2004). The link between subjective well-being and objective conditions in European societies. In W. Arts & L. Halman (Eds.), *European values at the turn of the millennium* (pp. 57-80). Boston: Brill.

Fajnzylber, P., Lederman, D., & Norman, L. (2002). Inequality and violent crime. *Journal of Law and Economics, 45*, 1-40.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science, 318*(5854), 1305-1308.

Gimpelson, V., & Treisman, D. (2015). *Misperceiving inequality* (No. w21174). National Bureau of Economic Research.

Gini, G. (1912). Variabilita e mutabilita contributo allo studio delle distributioni e delle relationi statistiche. *Universita degli Studi, Bologna*, *156*.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*(1), 1.

Goodwin, M. (2016, 28 June). *Inequality not Personalities Drove Britain to Brexit.* Retrieved from http://www.politico.eu/article/inequality-not-personalities-drove-britain-to-brexit/.

Grosfeld, I., & Senik, C. (2010). The emerging aversion to inequality: Evidence from subjective data. *Economics of Transition, 18*(1), 1-26.

Hagerty, M. R. (2000). Social comparisons of income in one's community: evidence from national surveys of income and happiness. *Journal of Personality and Social Psychology, 78*, 764-771.

Helliwell, J. F., & Huang, H. F. (2008). How's your government? International evidence linking good government and well-being. *British Journal of Political Science, 38*, 595-619.

Hirschman, A. O., & Rothschild, M. (1973). The changing tolerance for income inequality in the course of economic development with a mathematical appendix. *The Quarterly Journal of Economics, 87*(4), 544-566.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489-16493.

Kaltenthaler, K., Ceccoli, S., & Gelleny, R. (2008). Attitudes toward eliminating income inequality in Europe. *European Union Politics*, *9*(2), 217-241.

Kawachi, I., Kennedy, B. P., Lochner, K., & Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American Journal of Public Health, 87*, 1491-1498.

Kelley, J., & Evans, M. D. R. (2017a). Societal Inequality and individual subjective well-being: Results from 68 societies and over 200,000 individuals, 1981–2008. *Social Science Research*, *62,* 1-23.

Kelley, J., & Evans, M. D. R. (2017b). The new income inequality and well-being paradigm: Inequality has no effect on happiness in rich nations and normal times, varied effects in extraordinary circumstances, increases happiness in poor nations, and interacts with individuals' perceptions, attitudes, politics, and expectations for the future. *Social Science Research*, *62*, 39-74.

Kennedy, B. P., Kawachi, I., & Prothrow-Stith, D. (1996). Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. *British Medical Journal*, *312*, 1004-1007.

Kiatpongsan, S., & Norton, M. I. (2014). How much (more) should CEOs make? A universal desire for more equal pay. *Perspectives on Psychological Science*, *9*(6), 587-593.

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296-1310.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143-156.

Kraus, M. W., & Tan, J. J. (2015). Americans overestimate social class mobility. *Journal of Experimental Social Psychology*, *58,* 101-111.

Kreidl, M. (2000). Perceptions of poverty and wealth in western and post-communist countries. *Social Justice Research*, *13*(2), 151-176.

Krugman, Paul (2010): *Inequality and crises: coincidence or causation?* Presentation at the Luxembourg Income Study research conference: Inequality and the Status of the Middle Class: Lessons from the Luxembourg Income Study, Walferdange, June 28-30.

Kuhn, A. (2015). *The Individual Perception of Wage Inequality: A Measurement Framework and Some Empirical Evidence* (No. 9579). Institute for the Study of Labor (IZA).

Luttmer, E. F. P. (2005). Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics, 120*(3), 963-1002.

Lynch, J., Smith, G. D., Harper, S., Hillemeier, M., Ross, N., Kaplan, G. A., & Wolfson M. (2004). Is income inequality a determinant of population health? Part 1. A Systematic Review. *Milbank Quarterly, 82*, 5-99.

McCarty, N., Poole, K. T., & Rosenthal, H. (2006). *Polarized America: The dance of ideology and unequal riches.* (Vol. 5). Cambridge, Massachusetts: MIT Press.

Mellers, B. A. (1982). Equity judgment: A revision of Aristotelian views. *Journal of Experimental Psychology: General*, *111*(2), 242.

Melrose, K. L., Brown, G. D., & Wood, A. M. (2013). Am I abnormal? Relative rank and social norm effects in judgments of anxiety and depression symptom severity. *Journal of Behavioral Decision Making*, *26*(2), 174-184.

Ngamaba, K. H., Panagioti, M., & Armitage, C. J. (2017). Income inequality and subjective well-being: a systematic review and meta-analysis. *Quality of Life Research*, 1-20.

Niedrich, R. W., Sharma, S., & Wedell, D. H. (2001). Reference price and price perceptions: A comparison of alternative models. *Journal of Consumer Research*, *28*(3), 339-354.

Niehues, J. (2014). *Subjective perceptions of inequality and redistributive preferences: An international comparison.* IW-TRENDS discussion papers, (2).

Norton, M. I., & Ariely, D. (2011). Building a better America—One wealth quintile at a time. *Perspectives on Psychological Science*, *6*(1), 9-12.

Norton, M. I. (2013). All ranks are local: why humans are both (painfully) aware and (surprisingly) unaware of their lot in life. *Psychological Inquiry*, *24*(2), 124-125.

Norton, M. I. (2014). Unequality: who gets what and why it matters. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 151-155.

Norton, M. I., Neal, D. T., Govan, C. L., Ariely, D., & Holland, E. (2014). The not-so-common-wealth of Australia: Evidence for a cross-cultural desire for a more equal distribution of wealth. *Analyses of Social Issues and Public Policy*, *14*(1), 339-351.

O'Connell, M. (2004). Fairly satisfied: Economic equality, wealth and satisfaction. *Journal of Economic Psychology, 25*(3), 297-305.

OECD (2013), OECD Guidelines on Measuring Subjective Well-being, OECD Publishing. http://dx.doi.org/10.1787/9789264191655-en.

Oishi, S., Kesebir, S., & Diener, E. (2011). Income inequality and happiness. *Psychological Science, 22*(9), 1095-1100.

ONS UK (2012), "*Analysis of experimental subjective well-being data from the Annual Population Survey. April to September 2011*", Statistical release, 28 February, available online at: www.ons.gov.uk/ ons/rel/wellbeing/measuring-subjective-wellbeing-in-the-uk/analysis-of-experimental-subjective-wellbeing-data-from-the-annual-population-survey-april-september-2011/report-april-to-september-2011.html.

Oshio, T., & Kobayashi, M. (2011). Area-level income inequality and individual happiness: Evidence from Japan. *Journal of Happiness Studies*, *12*, 633-649.

Oswald, A. J., & Wu, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence from the USA. *Science*, *327*(5965), 576-579.

Ott, J. (2005). Level and inequality of happiness in nations: Does greater happiness of a greater number imply greater inequality in happiness? *Journal of Happiness Studies*, *6*(4), 397-420.

Page, L., & Goldstein, D. G. (2016). Subjective beliefs about the income distribution and preferences for redistribution. *Social Choice and Welfare*, 1-37.

Paskov, M., & Dewilde, C. (2012). Income inequality and solidarity in Europe. *Research in Social Stratification and Mobility, 30*, 415-432.

Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T., & Wilkinson, R. G. (2005). Wider income gaps, wider waistbands? An ecological study of obesity and income inequality. *Journal of Epidemiology and Community Health, 59*, 670-674.

Pickett, K. E., & Wilkinson, R. G. (2015). Income inequality and health: a causal review. *Social Science & Medicine*, *128,* 316-326.

Piketty, T., & Saez, E. (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, *118*(1), 1-41.

Pop, I. A., van Ingen, E., & van Oorschot, W. (2012). Inequality, wealth and health: Is decreasing income inequality the key to create healthier societies? *Social Indicators Research*, 1-19.

Sacks, D. W., Stevenson, B., & Wolfers, J. (2010). *Subjective well-being, income, economic development and growth* (No. w16441). National Bureau of Economic Research.

Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science, 308*(5722), 648-652.

Senik, C. (2004). When information dominates comparison: Learning from Russian subjective panel data. *Journal of Public Economics*, *88*, 2099-2123.

Solt, F. (2008). Economic inequality and democratic political engagement. *American Journal of Political Science, 52*(1), 48-60.

Stevenson, B., & Wolfers, J. (2008). Happiness inequality in the United States. *Journal of Legal Studies*, *37,* S33-S79, 352.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1-26.

Survey of Consumer Finance data (2016). Available online at: https://www.federalreserve.gov/econres/scfindex.htm and accessed using Damico, A. (2016). http://asdfree.com/survey-of-consumer-finances-scf.html.

Tomes, N. (1986). Income distribution, happiness and satisfaction: A direct test of the interdependent preferences model. *Journal of Economic Psychology*, *7*, 425-446.

Verme, P. (2011). Life satisfaction and income inequality. *Review of Income and Wealth*, *57*, 111-127.

Wilkinson, R., & Pickett, K. (2011). *The spirit level: Why greater equality makes societies stronger*. Bloomsbury Publishing USA.

Zheng, H. (2012). Do people die from income inequality of a decade ago?. *Social science & medicine, 75*(1), 36-45.

3.17 Appendices

3.17.1 Appendix I

Calculation of the Gini Coefficient

The Gini coefficient is defined as

$$G = \frac{MAD}{2\ MEAN}$$

where MAD is the mean absolute difference of the distribution. In the case of the lognormal distribution we have

$$G = \mathrm{erf}(\frac{\sigma}{2})$$

The mean of the lognormal is equal to $e^{\mu + \sigma^2/2}$ so we have that

$$MAD = \ 2e^{\mu + \sigma^2/2}\mathrm{erf}(\frac{\sigma}{2})$$

Because the offset lognormal just adds a constant amount to each individual’s income, the MAD is the same as the lognormal, since the differences between individual’s incomes do not change if they both get the same constant amount. The mean of the offset lognormal is

$$\theta + e^{\mu + \sigma^2/2}$$

where θ is the constant that we add. We thus find the Gini coefficient for the offset lognormal as

$$G = \frac{e^{\mu + \sigma^2/2}}{\theta + e^{\mu + \sigma^2/2}}\mathrm{erf}\left(\frac{\sigma}{2}\right)$$

Gini is bounded between zero and 1, with zero corresponding to a perfectly equal society, while 1 corresponds to a perfectly unequal society.

3.17.2 Appendix II

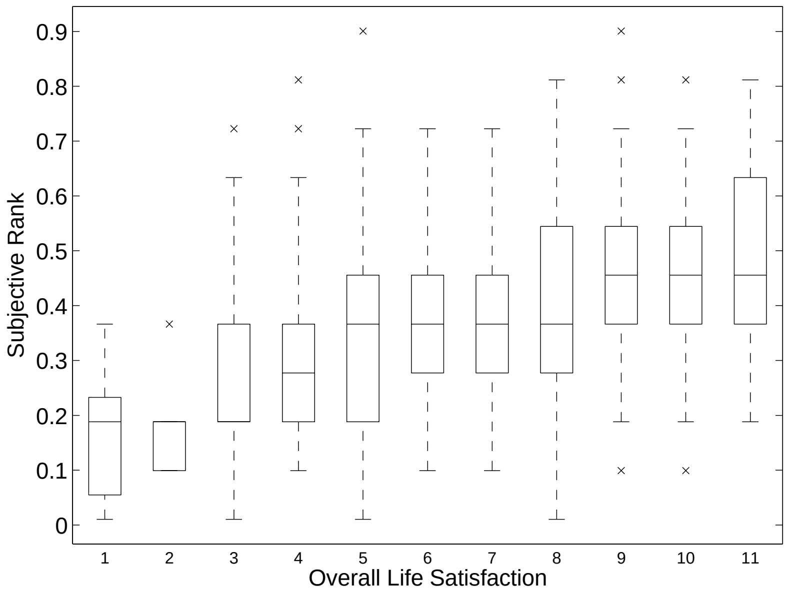


Figure U1. Relationship between Subjective Rank and Overall Life Satisfaction.

Table U1. OLS Regression results for measures of Well-being against Subjective Rank, Subjective Rank squared, Subjective Rank cubed, Age and Gender (simultaneous independent variables). DoF=545, p values in parentheses.

Well-being measure	Overall Life Satisfaction	Eudemonic Well-being	Happy Yesterday	Anxiety Yesterday	Ladder 1	Ladder 2
Subjective Rank	14.871781	7.64776	8.179383	- 11.1169	7.619413	14.952470
	(0.00225)	(0.1423)	(0.1059)	(0.06156)	(0.050149)	(0.000514)
Subjective Rank^2	-16.038482	-2.937651	-3.384331	13.90709	-0.814054	-22.973393
	(0.18602)	(0.8214)	(0.7888)	(0.34894)	(0.933169)	(0.0321780
Subjective Rank^3	5.418040	-2.192697	-3.306583	-3.81594	-1.717381	15.328619
	0.54886	(0.8213)	(0.7256)	(0.73027)	(0.812473)	(0.055166)
Age	0.009035	0.017187	0.008890	-0.02416	0.010054	-0.010742
	(0.22454)	(0.0317)	(0.2516)	(0.00818)	(0.091782)	(0.102219)
Gender	0.517519	0.496031	0.394638	-0.17646	0.482426	0.532070
	(0.00446)	(0.0111)	(0.0372)	(0.42692)	(0.000949)	(0.000944)
R^2 Multiple	0.1709	0.1094	0.09302	0.04918	0.2982	0.209
R^2 Adjusted	0.1633	0.1012	0.08469	0.04044	0.2918	0.2017
DoF	545	545	545	545	545	545

Table U2. OLS regression results. Dependent variable Delta Rank (Subjective-Objective).

	Coefficient	St. Error	*p* value
Intercept	-1.507e-01	2.420e-02	< .001
Personal Income	-1.132e-06	1.752e-07	< .001
Age	4.664e-04	4.248e-04	0.2728
Gender	-2.074e-02	9.977e-03	0.0381
Overall Conservatism	-1.547e-04	2.555e-04	0.5452
R^2 Multiple	0.07769		
R^2 Adjusted	0.07092		
DoF	545		

3.17.3 Appendix III

Table U3. OLS regression results. Did changes in Income, Overall Conservatism affect changes in Gini?

Dependent Variable Delta Gini	Coefficient	St. Error	*p* value
(Intercept)	-2.042e-02	4.807e-02	0.671
Delta Income	3.457e-07	5.076e-07	0.496
Age	4.577e-04	9.133e-04	0.617
Gender	6.103e-03	2.186e-02	0.780
Delta Overall Conservatism	-2.676e-05	1.222e-03	0.983
R^2 Multiple	0.002838		
R^2 Adjusted	-0.01141		
DoF	280		

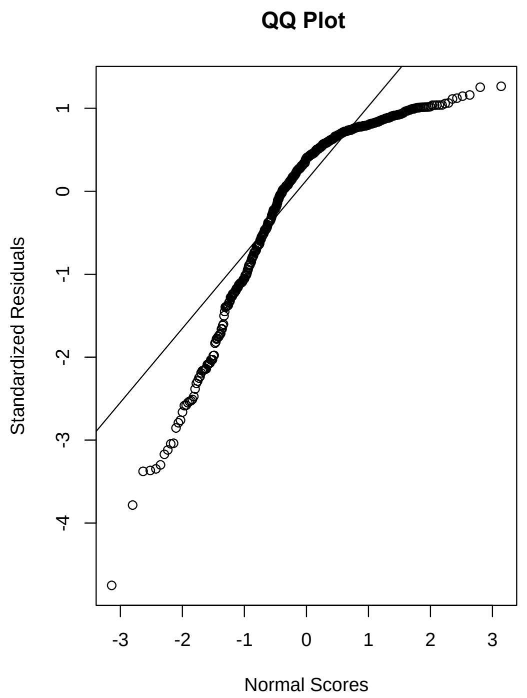
Table U4. OLS Regression results for measures of Delta Well-being against Delta Income, Subjective Rank etc. Independent variables were inserted individually in the regressions but we controlled for Age and Gender in all. Multiple R^2 reported, DoF=281, for all regressions.

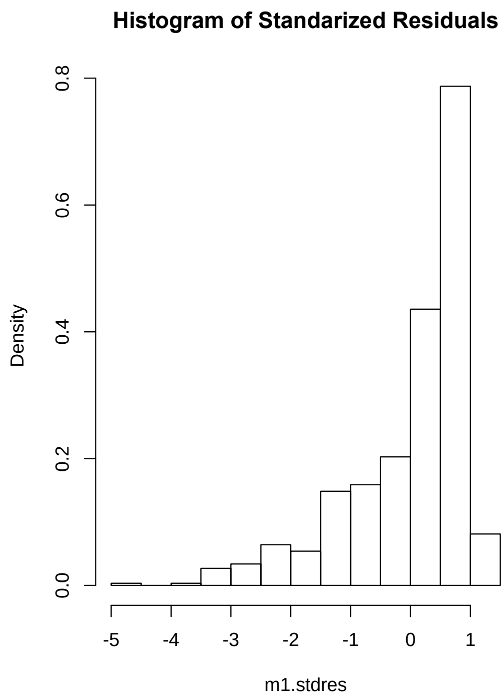
Wellbeing measure		Delta Income	Delta Log(1+Income)	Delta Subjective Rank	Delta Inferred Rank	Delta Objective Rank	Delta Gini
Delta Overall Life Satisfaction	Coefficient	3.541e-06	0.346813	1.815751	0.791262	1.305456	-0.110817
	p value	0.331	0.126	**0.0177**	0.157	0.0927	0.796
	R^2	0.007832	0.01274	0.02426	0.01156	0.01447	0.004719
Delta Eudemonic	Coefficient	8.065e-06	0.417693	0.575388	1.366677	1.077385	-0.499925
	p value	0.112	0.187	0.592	0.0796	0.321	0.403
	R^2	0.01352	0.01075	0.005619	0.01545	0.008091	0.007075
Delta Happy Yesterday	Coefficient	3.351e-06	0.314043	-0.349302	0.970636	1.279954	0.625086
	p value	0.506	0.317	0.743	0.209	0.234	`0.291
	R^2	0.001921	0.003904	0.0007301`	0.005944	0.005388	0.004306
Delta Anxiety Yesterday	Coefficient	-2.005e-06	-0.247869	-0.241723	-1.555905	-1.075770	-0.29771
	p value	0.760	0.545	0.862	0.123	0.443	0.700
	R^2	0.001032	0.002003	0.0008089	0.009152	0.002793	0.001228
Delta Ladder 1	Coefficient	5.019e-06	0.455089	2.123417	1.133265	1.735731	-0.350502
	p value	0.251	0.095**	**0.021**	0.0915	0.0628	0.496
	R^2	0.006663	0.01186	0.02076	0.01207	0.01424	0.003638
Delta Ladder 2	Coefficient	7.224e-06	0.499991	0.262910	0.815747	1.74020	-0.427786
	p value	0.0921	0.0614	0.772	0.216	0.0572	0.397
	R^2	0.01765	0.01997	0.007964	0.01306	0.02038	0.0102

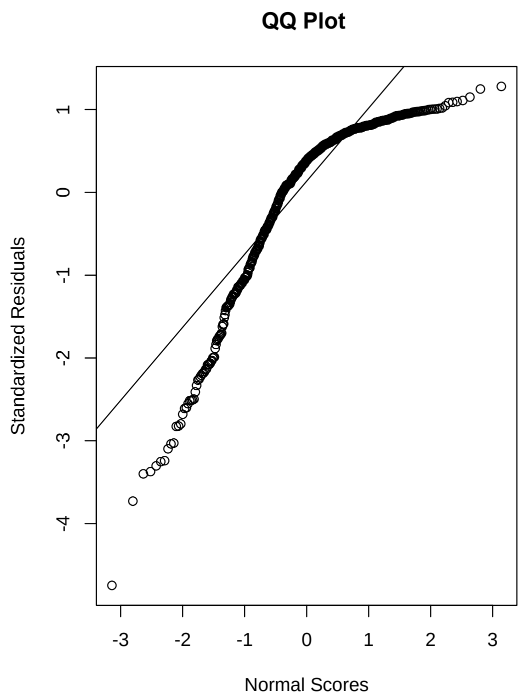
Delta denotes the difference between the values reported in Study 1 and Study 2. For example, Delta overall life satisfaction would be equal to the difference between the overall life satisfaction reported by an individual in Study 2 minus her overall life satisfaction reported in Study 1. We only observed a marginal effect for subjective rank; increasing subjective rank by 0.1 between the studies increased Delta overall life satisfaction by 0.18

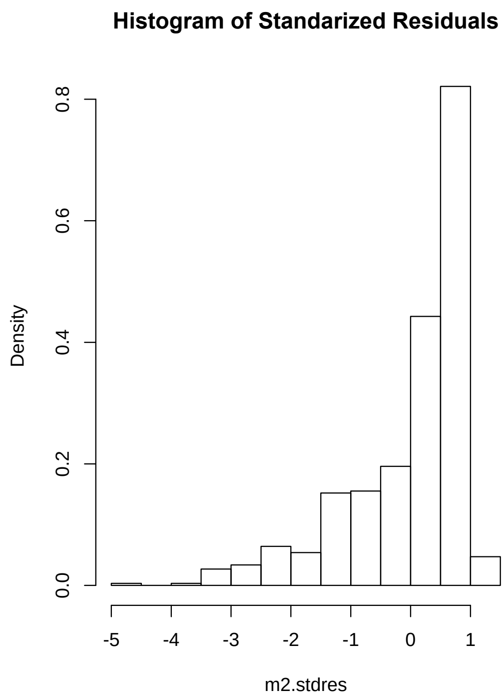
3.17.4 Appendix IV

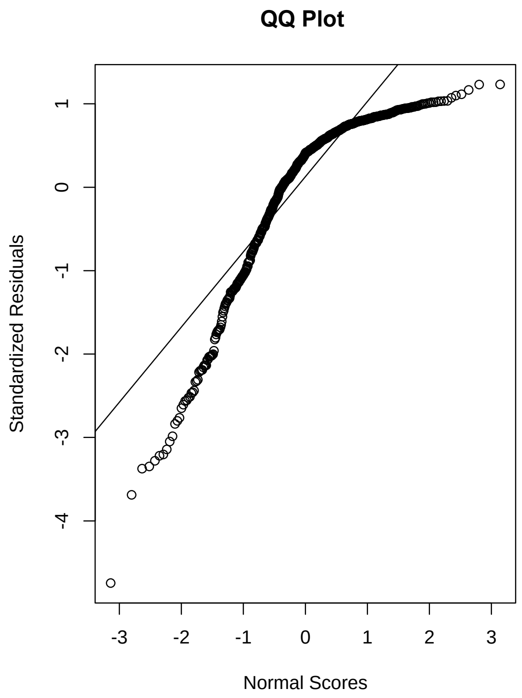
OLS Regression Diagnostics for Models 1 to 4

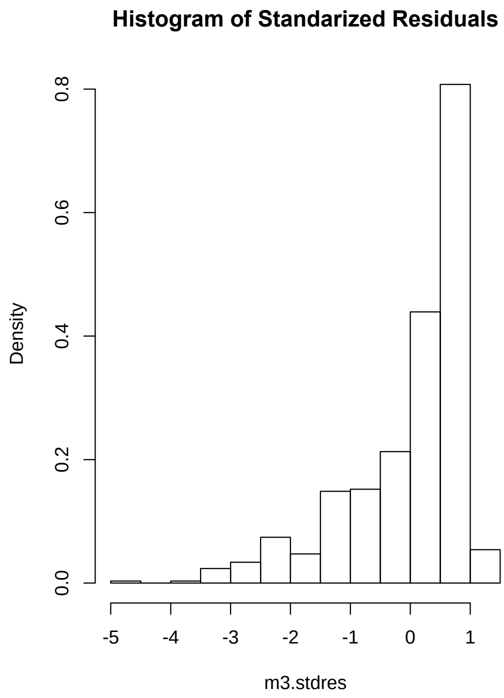


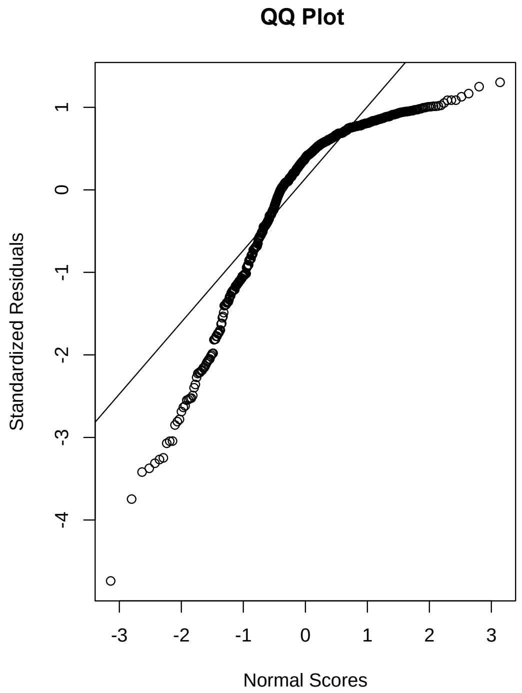


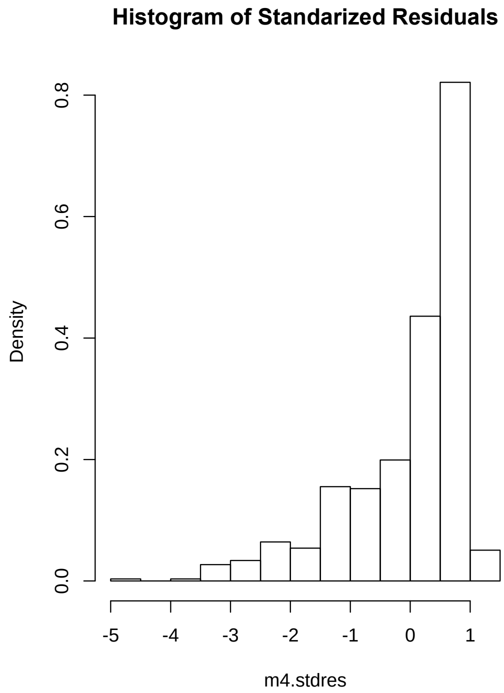












3.17.5 Appendix V

Notes on Quantile Regression

Quantile vs. Quintile

For a dataset $Y = \{y_1 \ldots y_n\}$ quintiles are rank values (from the Latin quinque for five), they are the (5-1) = 4 values $\{q_1, q_2, q_3 q_4\}$ such that (100/5) = 20% of the datapoints are less than q_1, 20% of the datapoints are between q_1 and q_2 and so on. In other words, they correspond to y values with rank positions 0.2, 0.4, 0.6 and 0.8 respectively.

Quantiles (from the Latin quantus meaning 'how much' or 'how many') can refer to any particular rank point in a dataset, and so quantile regression refers to a regression for particular rank points, like the median.

To connect them, quintiles are also known as 5-quantiles split it into 5 pieces = quintiles. In general, we can have n-quantiles for any n greater than or equal to 2. There are n-1 n-quantiles (there are four 5-quantiles) and in particular there is one 2-quantile which is the median.

OLS vs. Quantile Regression (Koenker & Hallock, 2001)

The mean of a dataset $Y = \{y_1 \ldots y_n\}$ is given by

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

But we can find the mean by minimizing a sum-of-squares. Define the loss function L_2 as a sum of squared differences of the data from a fixed value b.

$$L_2(b) = \sum_{i=1}^{n}(y_i - b)^2$$

Then we can minimize it by differentiating with respect to b

$$\frac{dL_2}{db} = \sum_{i=1}^{n} 2(y_i - b)$$

and solving

$$\frac{dL_2}{db} = 0$$

which gives us the equation

$$\sum_{i=1}^{n} 2(y_i - \bar{b}) = 0$$

which has the solution

$$\bar{b} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

i.e. we find the expected value of the data $E[y]$ because y is the random variable by minimizing the loss function L_2. We can also perform a linear regression in this way. In linear regression we minimize the sum-of-squares distances between y_i and the model

$$L_2(a, b) = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

The expected value conditioned on the value of x, is

$$E[Y|X = x]$$

which is the mean value of y given a particular value of x. Quantile regression is exactly the same, but instead of the mean, we use the median. We can calculate the median of a dataset by minimizing a loss function also (Baum, 2013).

$$L_1(b) = \sum_{i=1}^{n} |y_i - b|$$

Then the solution $\bar{b}$ is the median of the data Y. A simplified version of the proof is given by differentiating the function

$$\frac{dL_1}{db} = \sum_{i=1}^{n} \frac{d}{db}|y_i - b|$$

Now

$$\sum \frac{d}{db}|y_i - b| = \begin{cases} -1, & y_i > b \\ 1, & y_i < b \end{cases}$$

So we get

$$\frac{dL_1}{db} = -1 \times A + 1 \times B$$

where A is the number of data points greater than b and B is the number of data points less than b. This derivative is zero only if the number of datapoints less than b is equal to the number of data points greater than b. This is satisfied by the median.

Quantile regression (sometimes least absolute deviation) is simply performed by minimizing the absolute deviations as

$$L_1(a, b) = \sum_{i=1}^{n} |y_i - (a\, x_i + b)|$$

What it estimates is the conditional median, so it tells us what the median of *y* will be, given the value of *x*. If we interpret the results of OLS regression as giving the mean of the dependent variable conditioned on the value of the independent variable, LAD regression gives the median of the dependent variable conditioned on the value of the independent variable (Baum, 2013).

This means that the regression coefficients can be interpreted in almost the same way. In particular, suppose we have a quantile regression model $Median(y) = a\, x + b$. If we increase *x* by 1, the median of *y* values which have *x*=1 will increase by a. Suppose *x* is income, and *y* is Gini. Then *a**10,000 + *b* predicts the median Gini coefficient for all people with an income of $10,000. *a**11,000 + *b* predicts the median Gini coefficient for all people with an income of $11,000. So, if income increases by $1,000, then the median Gini coefficient changes by 1000*a.

Assumptions and Robustness

Quantile regression is more robust to unequal variances of the residuals, and skewness in residuals. In fact, if the residuals are highly skewed, then quantile regression is more efficient (meaning statistically powerful) than linear regression. One advantage of quantile regression, relative to the ordinary least squares regression, is that the quantile regression estimates are more robust against outliers in the response measurements.