Essays in Dishonesty

Submitted by Shaun Brian Grimshaw to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Economics
In September 2017

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:

Abstract

This thesis describes three different experiments investigating dishonesty. Chapter one investigates the use of default values and prompts in a tax filing system. Pre-populated fields simplify the process of filing taxes, thereby reducing the scope for errors. Such defaults may increase the scope for non-compliance if set incorrectly. The chapter describes an experiment investigating the effect of correct and incorrect defaults. The results show that setting defaults that underestimate taxpayers' true liability produces a fall in compliance. Nudges designed to mitigate the adverse effect of pre-population are also described. Nudges using descriptive norms in a dynamic manner that react to taxpayer decisions raise compliance. The chapter concludes that the use of defaults is worthwhile only if the data is of sufficient quality.

Chapter two describes a model for lying aversion containing cost elements in terms of the size of the lie told and in the positive deviation above a reference point reflecting the point at which someone becomes concerned about the credibility of the value being reported or about appearing boastful. An experiment based on a numeracy test where subjects have the ability to cheat by paying themselves for their performance is used to test the model. Two treatments are detailed using modal values from initial control sessions to set different reference points. The results show a greater propensity among subjects to report false values under the higher reference point consistent with the model.

Chapter three details an experimental investigation into lying behaviour between two samples, one a sample of undergraduate student subjects the other of workers recruited through Amazon Mechanical Turk. Results from a sender-receiver game based on a lottery draw show a higher propensity to report partially false values among student subjects, consistent with a higher reputational concern on behalf of the workers compared to students.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor, Miguel Fonseca, for his guidance, advice and dedication throughout the doctoral process. His ability to discern good ideas from bad and advice on a range of issues from from how to stage a laboratory through to how best to write to sell a story have been invaluable.

Thanks must also go to Todd Kaplan for guidance and advice on how to approach an economic model. My thanks also to Cecilia Chen for sharing her insight into experimental economics and assistance in various projects and to Luke and Kaustav for their insights into teaching.

My thanks to all in TARC, particularly Gareth, Lynn and Christos for inviting me to be part of the research centre, and to all those who have worked in or visited the offices, Ana, Diana, Hana Pol and Pei to name those who I spent the most time with there. Also to the researchers in HMRC who assisted with the work, most notably Joe, Rossy, Adrian and Soma. And a big thank you to Irene for her assistance in the co-ordination of any number of things.

I must also express my gratitude to members of the economics PGR group for interesting presentations and discussions. Too many to name, but particular thanks go to Greame, Lorenz, Hannes and Atisha who sat through many of the iterations of the thesis contents.

A big thank you must also go to Tim Miller, for guidance on technical issues and support in recruiting subjects, allowing for the experiments detailed in this thesis.

Finally, I would like to thank my family for their support. From listening to some of the more crazy ideas to double checking experimental materials, they have been there throughout and ensured everything stayed fun.

Introduction

Every day, people tell little lies. They tell little lies to themselves, believing their abilities to be greater than they truly are or that particular actions they take are not really harmful. People also tell little lies to others, often to spare someone's feelings or to avoid embarrassment. Some people also tell big lies. They rebuff having affairs to partners, deny crimes they have committed or simply misreport facts for personal gain. In a simple rational economic framework, people should lie so long as the benefits from the lie outweigh the costs. People are confronted with self-serving dishonest choices everyday, yet in contradiction to the simple theory honesty often prevails. It has been argued that a social norm of honesty exists because it allows for a level of trust between individuals and organisations that, without which, many of the mutually beneficial opportunities for cooperation would not be possible (Arrow 1970). In two key economics experiments, people have been shown to exhibit lying aversion, choosing different relative allocations when sending a false message is required compared to when the same choice is implemented directly (Gneezy 2005), and to not report the maximum possible value but to report partial lies, in so much as experimental subjects reported false values on unobserved dice rolls but did not always report the payment maximising value (Fischbacher and Föllmi-Heusi 2013). These results are consistent with the idea that people are not fully dishonest and that some degree of honesty, or at least a desire to appear honest does exist (Akerlof 1983).

The questions as to under what circumstances people are dishonest, and when they are dishonest what lies they tell, are therefore of particular interest. Such questions are not only of interest to academics, but also to policy makers. Dishonest actions may cost agencies directly, through theft or non-payment of monies due, but also indirectly through erosion of the level of trust alluded to in the previous paragraph. Information about when and how people are dishonest may assist in the creation of better policies and improved designs of systems and processes to reduce dishonesty and any associated negative impacts.

The following thesis experimentally examines dishonesty in three separate contexts. The first chapter investigates cheating in a tax-compliance setting, the second in self-reports of achievement on a numeracy test and the third in a game where one subject in an experiment must report the value they received in a lottery draw to another. The common feature of all three experiments is that, other than where the subject is already receiving the maximum, a false declaration can increase the level of payment. The chapters vary in the particular questions they attempt to address about dishonest behaviour.

The first chapter examines the effect on compliance of the use behavioural nudges in pre-populated tax forms, as recently published in Fonseca and Grimshaw (2017). The chapter describes a recent change in UK tax policy whereby individuals who reported their financial affairs to HMRC, the UK tax authority, through self-assessment are being moved to a new digital tax account system. A key innovation of the new system is that tax forms will be pre-populated with information held by the tax authority. This measure is designed to both reduce the level of complexity for the tax payer and the level of non-compliance through error (HMRC 2015a), but introduces a new problem in relation to the quality of the information held by the tax authority, primarily will the use of incorrect information by the tax authority affect compliance? The chapter investigates the effect on compliance of correct and incorrect pre-population, both in terms of to the tax-payers advantage in that earned income is missing from their account, and to their disadvantage, in that the pre-populated value is too high and would lead to an overpayment compared to the true tax due. Further to finding a reduction in compliance associated with the pre-population of tax forms in favour of the tax-payer, the chapter also investigates a number of behavioural prompts designed to restore compliance levels. Interestingly, the chapter finds that a lock designed to increase compliance in the case where the pre-populated values is correct further decreases compliance and that the use of a descriptive norm of behaviour was only found to be effective when used dynamically in response to a non-compliant filing. The chapter concludes with a warning to policy makers that pre-population should only be used where they are confident of the quality of the information being used.

The second chapter presents a psychological model of lying aversion based on a utility function with two cost terms, one to represent disutility in terms of the size of the lie, the second to represent disutility from reporting a value above a reference point. The reference point represents a value above which an agent in the model may feel more uncomfortable reporting an outcome, as they may perceive such a value to be less credible, to affect their perception of honesty or simply that the agent may wish to avoid bragging about their true state. The chapter details a number of propositions from the model in relation to how the value an agent will report changes with a change in the state and with the reference point. The chapter then describes an experimental test of the model using a numeracy test containing twenty questions where subjects were given the opportunity to cheat as they were required to pay themselves for the task and therefore self-report their outcome. Three treatments are described, a first to establish the basic parameters of the experiment and then two further treatments each with a separate reference point taken from the two modal values self-reported by subjects in the initial treatment. In one condition, LOW, subjects were informed that the modal value reported by previous subjects was ten. In the second condition, HIGH, subjects were informed of the other modal value, which was that subjects had reported twenty. The chapter details experimental results consistent with the model and reports on a variety of dishonest behaviours observed in the experiment.

A key contribution of the thesis is the novel design for the investigation of dishonesty presented in chapter 2. There are currently two major methodological strands in the economics literature for the investigation of lying. One is based on the sender-receiver game, as used in chapter 3, with its origins in the works of Gneezy (2005) and Sánchez-Páges and Vorsatz (2007), whereby the experimenter observes the true state along with sender subject and therefore can analyse individual lying. The second method uses unobserved responses, such as dice rolls or coin tosses as described in Fischbacher and Föllmi-Heusi (2013) and Abeler et al. (2014) which can be compared to statistical distributions or to observed responses for the same task, such as in Mazar et al. (2008). The approach used in chapter 2 is different in that subject responses are anonymous, but in having subjects hand in their test papers along with their returned coins, it is still possible to examine dishonesty at the individual level. The chapter describes that a considerable proportion of the false values reported, where subjects paid themselves for incorrectly answered questions, arouse from calculation error. It notes also, though, that in other cases the actual behaviour of subjects was less clear, in that the subjects made a number of mistakes that may have been calculation errors or may actually have been cheating. This design may offer an avenue for future research into dishonesty under the condition that the intention of the response is ambiguous.

The third chapter presents results from an on-line sender-receiver game undertaken by student subjects and by subjects recruited through the Amazon Mechanical Turk (AMT) system. Subjects were given the ability to increase both their own payment and that of a partner within the experiment by misrepresenting the value of a lottery draw. As well as examining for difference in behaviour between the two samples, the chapter describes two treatments, a first in which the lottery draw is uniform over the whole range, and a second treatment in which the distribution of the lottery is skewed towards the lower values. The chapter reports a higher tendency by student subjects to report self-serving false values than AMT subjects, in particular through false values that are not the maximum value possible. The chapter also reports that AMT subjects were observed to have a lower propensity to report the maximum value under the treatment where the value of lottery draw was skewed to lower values, consistent with the model of Chapter 2.

Chapter 1

Do behavioral nudges in pre-populated tax forms affect compliance? Experimental evidence with real taxpayers

1.1 Introduction

Governments are turning to the marketer's toolkit to design the way in which they interact with their citizens (Dolan et al. 2010; Thaler and Sunstein 2008). While nudges like default options were initially applied to consumer choice domains like insurance purchasing (Johnson et al. 1993), there has been an increasing recognition of the importance of the way choices are framed for nudging public policy relevant decisions, like organ donation (Johnson and Goldstein 2003) or retirement pension choices (Madrian and Shea 2001) - see Johnson et al. (2012) for a review.

In an important development in the public policy sphere, defaults are being introduced to the tax domain. In particular, the UK tax authority is now moving towards online tax filing (HMRC 2015a). Within that framework, it is using information about taxpayers' income and/or tax-deductible expenses from third parties such as employers, banks or pension companies to pre-populate the tax form. This move follows an international trend: the State of California already pre-populates elements of its state tax returns with the Ready Return program; tax return pre-population happens to varying degrees in over ten European Union countries, and Australia (European Commission 2012; OECD 2006). A pre-populated field in a tax form is effectively a default.

This chapter reports the results of an online experiment designed to understand the impact on filing behaviour of introducing defaults and norm-based nudges in online tax returns. The experiment contributes to the literature on default options by exploring the potential compliance benefits from pre-populating fields in tax returns; particularly, what the potential pitfalls are if default values are set incorrectly. The study also explores for the first time in a tax context, the potential for nudges that invoke descriptive norms to change compliance behaviour. The chapter considers static nudges, which have been the focus of attention in the social norm messaging literature. The chapter also looks at a novel form of nudges that react to users' inputs, which are especially well suited to online environments.

Taxes are an interesting, and relatively unexplored domain of research in marketing. They are a ubiquitous payment to all consumers, yet they differ from most personal consumer payments along two important dimensions, as noted by Lamberton (2013): most consumer purchases have an intrinsic personal benefit, and for the most part are something which consumers have control over. In contrast, taxes are compulsory payments, the benefits of which are not directly experienced by individuals, but rather experienced indirectly through the provision of public goods like roads, schools and law enforcement. This might explain why consumers often view taxes as a loss of personal freedom (Kirchler, 1998), and why consumers exhibit tax aversion, defined by Sussman and Olivola (2011) as "a dislike of taxes per se that goes above and beyond any associated financial costs."

Filing a tax return is similar to financial consumer decisions like choosing a pension plan. Both tasks are procedurally complex and cognitively demanding; they require a reasonable degree of financial literacy, as well as knowledge about the regulatory framework. Just as there is ample evidence of individuals and households making errors in their financial decision-making (Bernheim 1988; Beshears et al. 2013; Lusardi and Mitchell 2007, 2014), errors in filing decisions account for a significant portion of non-compliance: the UK government estimates it loses £6.5 billion (19%) of its tax revenues due to filing errors (National Audit Office 2015); Andreoni et al. (1988) estimate that 7% of US taxpayers make mistakes when filing their tax returns. Governments are starting to recognise the benefits from simplifying decision processes and helping people with their financial decision-making also apply to tax filing (Government Accountability Office (GAO) 2005; Reeson and Dunstall 2009). Moreover, the financial case for doing so is overwhelming.

That being said, tax filing differs from other types of financial decisions in important ways that are likely to have implications for the impact of defaults. When a company decides which of the 401(k) plans to introduce as a default, it knows its employees' income but not their preferences over plans. When a government agency pre-populates tax returns it assumes taxpayers wish to minimize their tax burden, but it only has potentially noisy third party data about taxpayers' true taxable income. It may have accurate salary data from employers, but only incomplete data on capital earnings, like dividend payments (Bloomquist et al. 2012). In that sense, the tax domain presents an important dimension for the use of nudges in public policy: honesty.

The asymmetry of information between the tax authority and taxpayers about the latter's income introduces a number of operational and ethical concerns when considering the introduction of pre-population. In particular, the tax authority could inadvertently pre-populate tax forms incorrectly. This may lead to a number of unintended consequences of a policy designed to improve compliance. One possible outcome is that the tax authority could underestimate taxpayers' liabilities. Taxpayers may simply accept the incorrectly pre-populated values because of status quo bias or behavioural inertia (Samuelson and Zeckhauser 1988), or because they trust the tax authority's assessment to be correct. A mistake of this nature would potentially leave the taxpayer open to an audit and any associated penalties from their non-compliance, since the legal responsibility for correctly filing the tax return still lies with the taxpayer. Another possible outcome is that the tax agency pre-populates tax returns in a way that over-estimates tax liabilities. Either over-estimating or under-estimating tax liabilities when pre-populating tax returns raises ethical considerations given the duty of care that tax administrations have towards taxpayers, and may lead to a public relations blow. It could also lead to additional audits being needed, the cost of which would offset the increase in overall revenue that pre-population offers.

The policy could make misreporting more prevalent by making non-compliance a passive act, rather than an active one (Mazar and Hawkins 2015; Spranca et al. 1991). Incorrect pre-population also reveals to taxpayers what the tax agency knows (and importantly, what it does not know) about their affairs; taxpayers may also interpret mistakes as incompetence on the part of the tax authority, thus extending the opportunity for deliberate evasion. In the context of defaults, the contribution to the literature is twofold. On one hand, we contribute to the literature on tax compliance by looking at the psychological determinants of compliance as a manifestation of honesty through the use of defaults. On the other hand, we also contribute to the understanding of acts of omission and acts of commission in the context of honesty.

Given the potential for unexpected non-compliance to emerge from the use of incorrect defaults, it is important to understand whether or not other types of nudges can be effective at mitigating any adverse effect of pre-population. Recently, Smith et al. (2013) advocated for the use of "smart nudges", which react to users' behaviour in real time. One of the benefits they propose is the potential to correct mistakes users may make along the decision process. Reactive nudges are of particular interest, as they are well suited to online environments such as the one considered by HMRC; we are also interested in understanding the extent to which they can mitigate potential errors coming from incorrect defaults. To this effect, we implemented a series of nudges, some of which included normative messages about compliance. Depending on the treatment, these messages appeared on screen as a function of the amounts declared by subjects on the tax form. To the best of our knowledge, this is the first study to test the effectiveness of reactive nudges on behaviour.

The following section develops the theoretical framework underpinning the experiment and the resulting hypotheses, followed by the experimental design. The chapter outlines the results and closes by discussing the relevance of the results to the academic and policy literatures.

1.2 Theoretical Framework and Hypotheses

If a tax authority pre-populates a particular entry on a tax form, such as employment income or taxable expenses, it is effectively imposing a default action on taxpayers. When deciding what value to enter as the default, tax agencies must consider that the information they have about that field may be potentially unreliable or uncertain. This introduces a moral dimension to the use of defaults, since taxpayers have an incentive to misrepresent their tax liabilities. Will defaults and other forms of nudges be effective in promoting desirable behaviour in this environment?

Gigerenzer (2010) argues that the same heuristics that guide choices in non-moral domains are also at play in moral decision-making. The default heuristic, which states that if there is a default, do nothing about it, has been put forward as the prime explanation for cross-country discrepancies in take-up rates of 401(k) savings plans with and without a default option (Madrian and Shea 2001) or sign up rates for organ donation (Johnson and Goldstein 2003), should also determine compliance behavior when filing a tax return with or without pre-population. Defaults should reduce the cognitive cost of making decisions, which should in turn help decision-makers understand more information and weigh information better (Peters et al. 2006), and ultimately reduce the time spent performing the task.

Kotakorpi and Laamanen (2015) study the effectiveness of pre-population on tax filing behavior by examining Finnish tax filing data. Between 1995 and 2004, Finland's tax authority started pre-populating sections of tax forms for a subset of the Finnish taxpayer population.

Kotakorpi and Laamanen (2015) report that taxpayers who received a pre-populated return were more likely to report the items that were pre-populated, and less likely to report on deductions that were not pre-populated. Kotakorpi and Laamanen (2015) only look at changes in filing behaviour in the income and deduction fields, and have no access to actual earnings data for those taxpayers. This is because the Finnish government did not audit any subset of taxpayers who were in the treatment and control groups. As such, their study cannot speak to whether pre-population led to changes in compliance. Importantly, a large proportion of taxpayers simply accepted the pre-populated returns, and chose not to file a modified form. The authors attribute this behaviour to taxpayers avoiding the cognitive and temporal costs associated with the complex process of engaging with (and potentially modifying) their tax forms.

Hypothesis 1 *Pre-populating income fields should lead to quicker completion of tax filing decisions across all treatments*

An important idiosyncrasy of tax filing decisions is that the tax authority may not always know taxpayers' true level of income. Any third-party information it resorts to in order to pre-populate a tax form could be incorrect. As such, introducing defaults in a tax context means it is possible to set the default value at an incorrect level: for instance, the tax authority could either overestimate or underestimate a taxpayer's income for that year.

Setting a default at a level that underestimates a taxpayer's income intrinsically changes the nature of a potential misrepresentation of income by the taxpayer from an act of commission to an act of omission. The literature on moral psychology has found consistent evidence for an omission bias in decision-making: morally reprehensible acts of commission are judged more harshly than acts of omission that carry the same negative consequence. Harmful acts of commission presumably signal malicious intentions on the part of the decision-maker, unlike acts of omission (Baron and Ritov 1994; Sánchez-Páges and Vorsatz 2009; Spranca et al. 1991).

The action principle in moral psychology (Cushman et al. 2006) also suggests that it is easier to passively refrain from acting morally than to actively transgress a moral norm. Teper and Inzlicht (2011) show that indeed this is the case for behavior in both prescriptive and proscriptive domains: participants in their experiment were more likely to offer to help a fellow student with a disability if they were asked directly than if they were passively given the option to help. Likewise, participants were less likely to cheat in a math quiz when doing so involved an action rather than an omission. Therefore it should be psychologically easier for participants to misreport their tax liabilities when doing so is a default action than when it is an active choice. (Duncan and Li 2017) report an increase in the proportion of payoff maximising dice rolls in an experiment where the payoff maximum value is pre-populated compared to no pre-population of the value. Finally, another potential motive for non-compliance following incorrect pre-population is behavioural inertia: taxpayers may simply accept the pre-populated values at face value and submit the tax return as it is (Madrian and Shea 2001; Samuelson and Zeckhauser 1988).

Hypothesis 2 *Pre-populating fields in a way that underestimates tax liabilities should lead to higher non-compliance*

Defaults can influence choice to the extent that decision-makers may believe they are a suggestion by the policy-maker, and as such imply a recommended course of action (Johnson and Goldstein 2003). In the tax context, they carry additional significance, as the tax authority, an expert body, and part of government, has a duty of care towards its taxpayers. If the tax authority chooses a default value which overestimates a taxpayer's liability, then taxpayers may interpret the incorrect pre-population as a signal that the tax authority does not have their interests at heart (Wright (2002) and Brown and Krishna (2004) define this process as "marketplace meta-cognition") or a signal of incompetence, and lead to higher non-compliance.

Hypothesis 3 *Pre-populating fields in a way that overestimates tax liabilities should lead to higher non-compliance.*

There are two approaches within the social sciences to conceptualize the determinants of honest behaviour. On the one hand, there is the external incentives approach, anchored in the economics of crime literature (Allingham and Sandmo 1972; Becker 1968). In this theoretical framework, the decision to be dishonest revolves around the calculus of expected utility: individuals weigh the relative gains from the dishonesty against the probability of being caught and the associated penalties. Therefore, increases in the expected benefit from dishonesty, either through changes in the penalties or probability of detection should increase the extent of dishonesty.

Independently to our own work, Bruner et al. (2015) study the role of defaults in taxes, based on the external incentives approach. The authors develop an individual tax evasion experiment, in which subjects are presented with a tax form in which, depending on the treatment, some fields are pre-populated. Subjects have two types of income sources: "matched" income, which is verifiable through third party data (e.g. salaries), and "unmatched" income (e.g. self-employment income). Their experimental design incorporates treatments in which itemized deductions are possible, to ascertain the impact of pre-population within the context of the US tax code. Their study looks at how changes in the regulatory framework, such as the audit likelihood or the presence of itemized deductions, affect behaviour with and without pre-population. Non-compliance is measured in three ways: underreported taxes on unmatched income, underreported taxes on the deduction, and overall underreported taxes. The authors find that pre-populating tax returns in a way that implies a lower tax liability increases non-compliance.

On the other hand, there is the internal incentives approach, rooted in the social psychology literature. This theoretical framework is based on the idea that moral actions are driven by internal rewards, which are, to some extent, uncorrelated with the degree of financial reward at stake. People develop an understanding what socially normative behaviour is from an early age (Campbell 1964), as well as through group membership (Akerlof and Kranton 2002; Sherif and Sherif 1953). Norms in turn establish a set of prescribed behaviours from which we derive psychologically wellbeing. Neuroeconomics research supports this account: individuals who cooperate or enforce cooperative behaviour in social dilemmas exhibit similar brain activation patterns (De Quervain et al. 2004; Rilling et al. 2002) as those observed when people experience financial rewards or when people consume food (Knutson et al. 2001; O'Doherty et al. 2002).

A particularly useful conceptualization of how internal reward mechanisms determine honest behaviour is given by the theory of self-concept maintenance developed by Mazar and Ariely (2006) and Mazar et al. (2008). It postulates that individuals, when considering whether to break a social norm such as honesty, trade-off the financial gains from cheating against maintaining a positive self-concept as an honest person. According to this theory, individuals manage this problem by finding a balance between these competing psychological demands. They do so by engaging in an amount of dishonest behaviour which brings in financial rewards but which is not enough to force people to reassess their self-image.

In this context, descriptive norms, norms pertaining to what people in a group or population do, can be a powerful driver of self-concept maintenance by prescribing a particular mode of behaviour (Cialdini et al. 2006). Descriptive norms can be a powerful driver of behaviour, ranging from littering (Cialdini et al. 1990; Reno et al. 1993) , environmentally-friendly behaviour (Goldstein et al. 2008), or energy consumption in the household (Schultz et al. 2007).

Hypothesis 4 *Introducing nudges that remind participants of descriptively normative behaviour should reduce non-compliance.*

Importantly, when developing the Focus Theory of Normative Conduct, Cialdini et al. (1990) and Cialdini and Reno (1991) argue that the effectiveness of social norm messages is critically dependent on whether they are focal in the decision-maker's attention, and therefore consciously salient. This approach suggests that nudges containing descriptive norm information will be more effective at preventing non-compliance if they are reactive to users' behaviour. Furthermore, the fact that the message displayed by the computer reacts to user behaviour should create a perception among users that the message is directly targeted at them. This "personalization" of the message content should increase the effectiveness of the reactive nudge, as per the literature on survey responses (Kanuk and Berenson 1975; Yu and Cooper 1983), as opposed to static nudges.

Hypothesis 5 *Nudges that are activated by non-compliant filing behavior should lead to more compliance than always-present static nudges.*

1.3 Materials and Methods

1.3.1 The Experimental Task

Participants in our study took on the role of a fictitious taxpayer. Their task was to complete a tax form based on a profile of income and expenses for that fictitious taxpayer. The profile detailed two sources of income and two corresponding expenses that could be used to reduce tax liabilities. Table 1.1 outlines the profile used in the experiment. Payoffs were denominated in Experimental Currency Units (ECU); 1,000 ECU were worth £0.50 (at the time, $0.75).

Field	Description	Value (in ECU)
Self Employment Income	Income from contract with local authority	25,200
Self Employment Income	Income for work done for ACS Ltd	27,100
Self Employment Expenses	Cost of Travel to Work	2,500
Property Income	Revenue from letting a flat	20,000
Property Expenses	Cost of estate agent and legal fees for letting of flat	2000 times the roll of a 6-sided die

Table 1.1: Contents of the taxpayer profile used in the experiment

The experimental instructions (see the Appendix for a copy of the instructions) detailed that participants would be paid according to the income in their profile minus any tax or fines due from their tax declaration and any potential audit. The instructions also detailed that upon filing their tax return the "experimental tax authority" could audit their tax return. If a participant's tax return was audited, the computer compared the values in the tax return with the values in the profile. The probability with which the experimental tax authority carried out audits was a function of the actual declared tax liability on the return, but it could never exceed 10%: the probability of audit was p=3.3% if the declared liability was greater than or equal to 45,200 ECU, p=6.6% if the declared liability was between 22,600 ECU and 45,199 ECU and p=10% if the declared liability was less than 22,600 ECU. Subjects did not know the actual probabilities of audit or how it changed as a function of declared tax liabilities. Subjects only knew that the probability varied with the amount declared and was limited to 10%. The chosen values did not intend to mimic the audit policy of the IRS, HMRC or any other tax authority.

Participants were required to submit a tax return based on the following fixed (and known) parameters: a tax rate of 40% and a penalty rate applied to unpaid tax of 50%. The values for the probability of audit, the tax rate and the fine rate were set so that the optimal action for a risk-neutral, payoff maximising participant was to under-report their tax liability - this matches the reality in the field (Allingham and Sandmo 1972; Andreoni et al. 1988).

Although the instructions did not instruct them to do so, participants could increase their financial payment by evading. They could do so by under-declaring income or by over-declaring expenses. In either case, the most they could gain would be to declare a tax liability of zero. This translates into a possible gain relative to full compliance of £13.56 (at the time, US$20.34) for a task that took on average 22 minutes[1].

After reading the instructions, participants were asked to complete a practice tax form based on a simple profile for which they were told they would not be paid. Upon completion of the practice form, participants were informed of what payoffs they would have received had they been audited or had they not been audited on

[1] Broken down as an average of seven minutes to read the instructions, two minutes to perform the practice round, three minutes for the tax filing and 10 minutes to complete the questionnaire.

their practice tax declaration.

The majority of items in the tax return were verifiable if audited. Verifiability is essential for income amounts, as these form a direct part of the participant's payoff; as such the experimenter must know the value in order to be able to pay it. Expenses, however, offer the experimenter the ability to set unverifiable items, in that the expenses act to reduce the tax paid, so participants can increase their payoff by raising expenses, but the experimenter does not need to know the true value. Unverifiable expenses potentially allow subjects a greater opportunity to evade, a mechanism found to have an effect in empirical studies (Kleven et al. 2012). The value of one of the expenses (i.e. Property Expenses) was allocated to be equal to the roll of a six-sided die multiplied by 2000 Experimental Currency Units (ECU). As a participant's dice roll is unverifiable, it is rational for them to declare the maximum allowable value for the expenses field - that is, 12,000 ECU, equal to rolling a six. While it can never be verified as to whether an individual misreported that expense item, non-compliance can be detected at the sample level, since the distribution of die rolls (and therefore of declared values on that item in the tax return) should be uniform if subjects are compliant (Fischbacher and Föllmi-Heusi 2013).

Once participants had completed their tax form, they saw their tax calculation on the screen. This was done to remove any computational burden from the participants, and it is similar to tax calculators which are available online. They could then either repeat the process in order to change their details or submit their tax return. After submission, the computer randomly determined if they were to be audited and the participants saw their payoff from the experiment.

Subjects then completed a questionnaire; it included two open-ended questions about their choices in the experiment, questions regarding participants' attitudes towards tax using Likert scales, as well as socio-demographic characteristics. They were informed that the questionnaire would not impact their payoff and they were able to leave any question blank if they wished. Finally, participants were told they had completed the experiment and given details of how to opt out of having their responses included in the data set, had they wished to do so.

A participant's experimental balance was calculated at the end of the experiment as the total of the two income streams in the profile minus the tax payable on their declared liability and any fines occurred from the under-payment of tax due. It is important to note that over-declaration of income could not raise participants' payoffs, and the experimental instructions were clear about this. Participants' earnings in ECU were converted to cash at a rate of 50p per 1,000 ECU; average earnings were £29.62 (at the time US$ 44.43).

1.3.2 Experimental Design

Treatment	Description
BASE	No information reported and all four fields left blank
CORR	Correct self-employment income streams reported, correct self-employment income pre-populated
OVER	Double counting of one income stream reported, incorrect (value too high) self-employment income pre-populated
UNDER	Omission of one income stream reported, incorrect (value too low) self-employment income pre-populated
UNDERGENERIC	Omission of one income stream reported, incorrect (value too low) self-employment income pre-populated, click of checkbox required to edit pre-populated field (and confirmation of edit)
UNDERALWAYS	Omission of one income stream reported, incorrect (value too low) self-employment income pre-populated. Additional message on screen: "Most people in your circumstances enter an income value of more than 40,000. Values below this amount are more likely to be audited. Click the tickbox to confirm you wish to proceed."
UNDERTRIGGER	Omission of one income stream reported, incorrect (value too low) self-employment income pre-populated. Same message as UNDERALWAYS only displayed if subject files self-employment income value less than 40,000.

Table 1.2: Treatments used in the experiment

The experiment consists of seven different treatments in a between-subjects design, summarised in Table 1.2. In the baseline treatment, BASE, the tax form was not pre-populated. In the CORR treatment, the tax form had the self-employment income field pre-populated with the same total amount as in the profile, the sum of the two values given for self-employment income, and the tax form displayed that the information held in the tax authority database was the two values corresponding to the two self-employment income streams in the profile. This corresponds to the case where the tax authority has access to quality third party reporting and therefore can correctly pre-populate the taxpayer's income (Gale and Holtzblatt 1997). In the UK, third party reporting forms the basis of the Pay-As-You-Earn (PAYE) system, such that the correct tax is paid at source and many employees are not required to submit a year-end tax return.

In the UNDER treatment, the self-employment income field was pre-populated with an incorrect value equal to one of the two sub-items of the self-employment income in the profile and the tax form displayed that the information held in the tax authority database was that single income stream. This captures the case where the tax authority has either incomplete access to third-party data (e.g. an employer not providing this information), or the case where the tax authority is unaware of that stream of income. This error in pre-population leads the tax authority to under-estimate the tax liability of the subject. In the OVER treatment, the tax form displayed that the information held in the tax authority database consisted of three values, where one of these was a double-counted entry. Hence, the value used to pre-populate the self-employment field of the tax form was greater than the actual income level in the subject's profile. This error in pre-population leads the tax authority to over-estimate the tax liability of the subject.

As argued in H2, we expected a large incidence of non-compliance in the UNDER treatment, either because inertia leads subjects not to change their pre-populated entries, or because subjects learn of the experimental tax authority's ignorance of the true profile values, and engage in active non-compliance. To test whether behavioural nudges can mitigate the negative effects of incorrect pre-population, we consider three additional versions of the UNDER treatment. The first was UNDERGENERIC, in which the pre-populated value was locked. In order to edit that field, participants had to first click a checkbox positioned next to it. In addition, participants also had to re-check that box in order to confirm the new value they inputted before filing the tax return.

The second version was UNDERALWAYS, which featured the following message: “Most people in your circumstances enter an income value of more than 40,000. Values below this amount are more likely to be audited. Click the tickbox to confirm you wish to proceed.” This treatment was intended to trigger a descriptive norm of compliance and reminded subjects of the nature of the audit rule. Social psychologists have long argued for the effectiveness of descriptive norms as catalysts of behaviour change, e.g. Cialdini et al. (2006). See Onu and Oats (2014) for a review of the evidence of norms applied to tax compliance.

Recently, a case has been made for the inclusion of reactive defaults. Smith et al. (2013) propose using defaults that react to inputs by decision-makers. We implement a treatment that approximates this recommendation: UNDERTRIGGER, in which the same message as UNDERALWAYS was featured, but only if the participant inputted a total self-employment income amount lower than 40,000. That is the same descriptive norm is used in both the UNDERALWAYS and UNDERTRIGGER treatments and in the UNDERALWAYS treatment the message is displayed immediately upon the load of the tax form, whereas in the UNDERTRIGGER treatment the message is only displayed after a subject has made their submission if the level of self-employment income reported is less than 40,000. In both cases the message must be acknowledge to allow the submission of a value of less than 40,000.

Our choice of nudge in the UNDERALWAYS and UNDERTRIGGER treatments was based on one of the mechanisms used by tax authorities to identify tax evaders, which is to target outliers from within a given group, for instance based on industry. For example the “DIF score” of the IRS in the USA will produce “audit flags” for taxpayers deviate from the average behaviour of their group (Alm and McKee 2004). As the probability of audit is endogenous with respect to the subject’s declaration in the experiment, we can use a nudge to inform subjects of the tax authority's operational process. We opted for the value for income displayed in the nudge to be below the actual value given in the profile, reflecting the process whereby outlying declarations are subject to higher probability of audit. It was also chosen to be above the value used for the pre-population in order for the message to have some degree of saliency.

1.3.3 Sample and Recruitment

Our sample consisted of a pool of participants who volunteer to take part in telephone and online surveys run by ICM, a market research company. ICM sent an invitation email to its participant pool to take part in an online decision-making experiment. As part of registering their interest to take part, participants were asked to fill in a questionnaire, consisting of a series of standard demographic questions. ICM screened participants on the basis of them stating to be over 18 years old, and either self-employed, or in full-time employment; this meant they were UK tax residents. ICM then invited at random 755 people from those who met our sampling criteria. Participants were required to have access to the web, as well as a six-sided die; the invitation email included a number of online links for simulated dice roll web sites for those that did not have access to a physical die. Out of the 755 individuals invited, 554 completed the experiment[2]. Just over 60% of our participants were male; participants' age ranged from 18 to 78; the average age was 44.5 years for women and 47.3 for men. 34% of subjects were self-employed and 66% were employees - the gender distribution was roughly the same for both employment categories.

1.3.4 Experimental Procedures

The experiment was operationalized through a customized website designed by the experimenters and hosted by the University of Exeter. The experiment took

[2]The dropout rates of those who started the experiment but failed to complete it were consistent between the treatments. There was, some variation in the numbers completing the experiment for each treatment, detailed in Table A.1 in the Appendix. The differences in the number of subjects arose from different proportions of those invited by ICM who accessed the experiment

place between 9 February and 12 April 2015. ICM provided each participant with a link to the experimental website and a unique login username and password. We could not match usernames to actual participant data, and ICM did not have access to participant decisions, making this a double blind experimental design. This was made explicit to participants when they were invited to participate.

Upon login each participant read an on-screen set of instructions that detailed the task they were required to perform. Participants were also told they would be paid a fixed £5 (US$ 7.50) for completion of the experiment and would have the opportunity to earn more based on their decisions in the experiment. The instructions detailed a number of examples of the potential outcomes from various declaration choices - the full set of screenshots is in the Appendix.

1.4 Results

We treat each individual decision in our analysis as an independent observation and make treatment comparisons using standard statistical tests. Unless stated otherwise, we report two-sided tests throughout. We complement these with econometric analysis, which also incorporates individual characteristics, as well as responses to the post-experimental questionnaire.

The major part of the analysis will be based on the fields in the tax form that are verifiable by the experimenters - therefore excluding Property Expenses. Observable Tax Liability is defined in the following manner:

$$\begin{aligned}\text{ObservableTaxLiability} = \text{SelfEmploymentIncome} + \text{PropertyIncome} \\ - \text{SelfEmploymentExpenses}\end{aligned} \tag{1.1}$$

Based on this variable, Compliance is defined as:

$$Compliance = \frac{(DeclaredObservableTaxLiability)}{(ActualObservableTaxLiability)} \tag{1.2}$$

If an individual has a Compliance value of 1, then that he/she is classified as

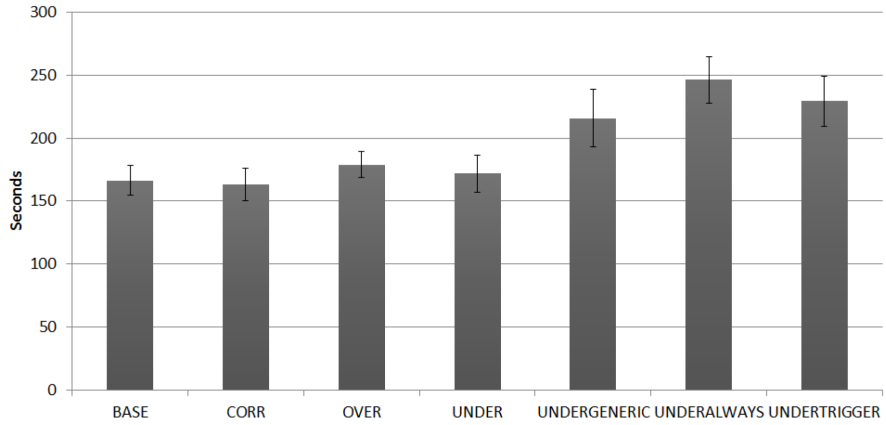


Figure 1.1: Average completion time for the experimental task

Compliant. This ratio defines two types of non-compliance. If the ratio is smaller than 1, that individual is Under-Compliant; if instead that ratio is higher than 1, that individual is Over-Compliant. The last type is relevant because the pre-populated amount in the OVER condition over-estimated the subject's taxable income, and passive acceptance of the default value could lead to subjects overpaying taxes.

1.4.1 Pre-Population and the Default Heuristic

To address the first hypothesis we analyze data on completion times for the tax-filing component of the experiment. The median completion time in our sample was 155 seconds, and 90% of subjects completed the filing task in 6 minutes or less. We do not include in the analysis of completion times four outlier individuals who took more than 90 minutes to finish.

If subjects passively accept defaults as part of their filing process, as the heuristics approach would suggest, we should observe shorter completion times in the treatments with defaults than the BASE treatment. Figure 1.1 displays average completion times across all seven treatments. Compared to BASE, we either found no significant difference in completion times (CORR: z=.65, p=.514; UNDERGENERIC (Mann-Whitney test, (henceforth MW) z = 1.73, p = 0.083), or there were longer completion times than in BASE: UNDERALWAYS (MW: z = 4.14, $p < .001$) and UNDERTRIGGER (MW: z = 2.83, p = 0.005).

Result 1 *The introduction of defaults does not lead to shorter average completion*

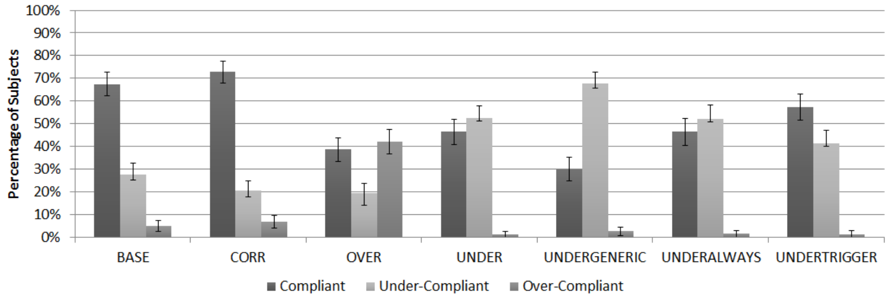


Figure 1.2: Percentage of Compliant, Under-Compliant and Over-Compliant subjects

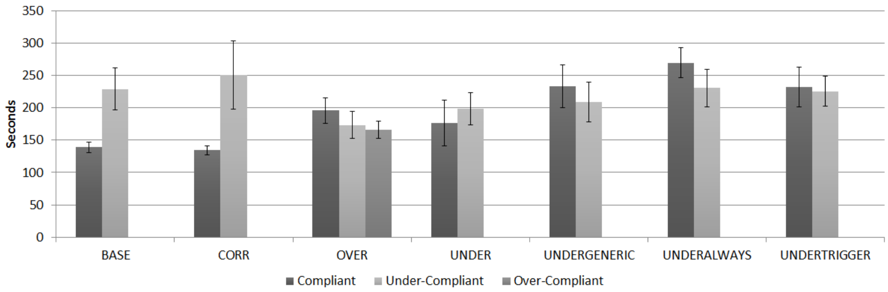


Figure 1.3: Average completion time for the experimental task conditional on subject type

times.

This suggests that, at least for some subjects, the presence of defaults led to greater deliberation with regards to what amounts to declare. As such, it is relevant to condition our analysis on compliance behavior.

Figure 1.2 shows the proportion of Compliant, Under-Compliant and Over-Compliant subjects in each treatment; Figure 1.3 outlines the average completion time for each type of subject. We note that in all but the OVER treatment, the number of Over-Compliant subjects is extremely small, so there is no meaningful analysis to be done in those cases, so we only report completion times for Over-Compliant subjects in the OVER treatment.

It is noteworthy that in the BASE treatment, Under-Compliant types took on average 90 seconds longer to complete the tax return than Compliant types (MW: $z = 3.05$, $p = 0.002$). The same is true in CORR, where the correct amount was already pre-populated; in fact, there is no statistically significant difference in completion times between CORR and BASE for either Compliant (MW: $z = .351$, $p = 0.726$) or Under-Compliant types (MW: $z = 0.126$, $p = 0.900$).

We note that Compliant individuals take longer to file their return in OVER than in BASE (MW: z = 2.497, p = 0.013) and CORR (MW: z = 2.654, p = 0.008). In contrast, Under-Compliant types take as long in OVER as in BASE (MW: z = 1.416, p = 0.157) or CORR (MW: z = 1.187, p = 0.253). Interestingly, the 40% of Over-Compliant individuals in OVER took as long as the Compliant (MW: z = 0.835, p = 0.404) and Under-Compliant types (MW: z = 0.680, p = 0.497) to complete the tax return.

Result 2 *Under-compliance requires greater deliberation time than compliance when those behaviors are acts of commission.*

The average completion times by Compliant individuals in UNDER and OVER are not significantly different to their counterparts in BASE and CORR. The same is true for Under-Compliant participants. However, when we break up the group of Over-Compliant in OVER into those who did not alter the pre-populated amount (N=28) and those who did (N=7), the average completion time of the former subgroup is 139 seconds, while the completion time of the latter is 274 seconds. The average completion time of those who accepted the incorrect default is not significantly different to the completion time by Compliant types in CORR (MW: z = 0.142, p = 0.887). This suggests that the behavior of some subjects may be driven by inertia.

Likewise, if we split the sample of Under-Compliant individuals in UNDER into those who left the default value unchanged (N=27) and those who did not (N=16), a similar pattern emerges to that observed in OVER. Those who accepted the default value and under-reported their tax liabilities took on average 142 seconds to complete the filing task; this is not significantly different to the average completion time by Compliant types in BASE (MW z = 0.801, p = 0.423) or CORR (MW: z = 0.672, p = 0.502). This suggests that the behavior of some subjects in both treatments is also driven by the default heuristic.

We can therefore classify Under-Compliant individuals in UNDER in two categories: 15 subjects (or 35%) are passively under-compliant, who accept the default value and are honest otherwise. The remainder are actively under-compliant:

some keep the incorrect default value and evade also in other fields in the tax return; others change all fields, including the pre-populated one.

Result 3 *A minority of individuals passively accepts incorrect defaults; their compliance behavior is primarily a function of the pre-populated value. However, most under-compliance is driven by active choice.*

In short, our data broadly rejects Hypothesis 1. Only a minority of subjects' behavior is consistent with the default heuristic. Most deviations in completion time are explained by deliberate actions.

1.4.2 The Effect of Pre-Populating Tax Forms

We start by looking at the impact of correctly pre-populating income fields in tax forms on compliance. Figure 1.2 indicates that in the BASE, just over two thirds of subjects were Compliant, and 28% were Under-Compliant. There are relatively more Compliant subjects (73%) and relatively fewer Under-Compliant subjects (20%) in CORR, although the distributions of types are not significantly different in the two treatments (Fisher's exact test (henceforth FET) $p = 0.574$). However, the proportion of subjects entering the correct value in the income field that is subject to pre-population went from 75% in the BASE to 99% in CORR (MW: $z = 4.356$, $p < .001$).

Result 4 *Correctly pre-populating a field in the tax form leads to higher compliance in that field, although not to higher compliance overall.*

In contrast, pre-populating the income field incorrectly lead to a decrease in the proportion of Compliant subjects. In the OVER treatment that is due to an increase in the proportion of Over-Compliant individuals, while the proportion of Under-Compliant subjects is unchanged. As a result, we observe a significant difference in the distributions of types in BASE and OVER (FET: $p < 0.001$). In the UNDER treatment, the drop in Compliant individuals relative to BASE is due to an increase in Under-Compliant types, while the proportion of Over-Compliant types remains unchanged. The type distributions in BASE and UNDER are significantly different (FET: $p = 0.002$).

Result 5 *Introducing incorrectly pre-populated fields results in fewer compliant subjects.*

In short, our data provide strong support for Hypotheses 2 and 3.

1.4.3 The Effectiveness of Norm-Conveying Nudges

The next question is to what extent can nudges imbedded within the tax return mitigate the adverse effects of incorrect pre-population. To do this, we use the UNDER treatment as the de facto baseline condition, and see whether nudges can "recover" compliance levels back to those observed in the original BASE treatment (or even higher).

Figure 1.2 shows that the effect of nudges on compliance is rather mixed: in UNDERGENERIC (which featured a checkbox that subjects had to un-tick before altering the content of the pre-populated field), the proportion of fully compliant subjects is significantly lower than in the UNDER treatment (FET: $p = 0.049$). Figure 1.3 shows that while average completion time for Compliant types was significantly higher in UNDERGENERIC than UNDER (MW: $z = 3.103$, $p = 0.002$), there was no significantly difference in average completion time among Under-Compliant types in both treatments (MW: $z = .030$, $p = 0.976$). In other words, the introduction of a check box increased under-compliance by introducing a physical barrier to changing the pre-populated field. This is manifested in the extra 56 seconds it took Compliant subjects to file their returns in UNDERGENERIC compared to Compliant subjects in UNDER.

The introduction of a descriptive norm message plus a confirmation tick box (UNDERALWAYS) had no discernible effect on the proportion of compliant individuals (FET: $p = 1.000$). It increased the average length of time Compliant subjects took to complete the tax return by over 90 seconds (MW: $z = 4.592$, $p < 0.001$) compared to UNDER. There was no significant difference in the completion time of Under-Compliant subjects in UNDERALWAYS (MW: $z = 1.555$, $p = 0.120$).

Passive under-compliance, as described in relation to the UNDER treatment above, was more difficult in the UNDERALWAYS treatment, as subjects were forced to acknowledge the statement of the norm by checking a tick box before they were able to submit their return. In other words, despite creating a psychological barrier to incorrect filing of the tax return, which manifested itself in longer average completion times for the Compliant subjects, there was little effect in terms of dissuading under-compliance.

Finally, the same message when triggered by subject's filing behavior (UNDERTRIGGER) led to a greater proportion of Compliant individuals than Under-Compliant in that treatment, although the distributions of subject types in UNDER and UNDERTRIGGER were not significantly different (FET: $p = 0.200$). Again, the effect of the norm trigger is felt on the Compliant types' behavior: they took on average 45 seconds longer in UNDERTRIGGER than in UNDER (MW: $z = 2.446$, $p = .014$).

Result 6 *Nudges based on physical barriers to changing default entries in tax forms compounded the under-compliance that exists with incorrect pre-population. Messages with descriptive norms did not achieve significant increases in the proportion of Compliant types.*

1.4.4 Observable Characteristics and Self-Reported Measures

As part of the post-experimental questionnaire, we collected data on a number of socio-demographic variables: age, gender, employment status and self-reported annual income. We also asked subjects a question related to their attitudes towards tax in general: "Do you think cheating on taxes if you have a chance is justifiable? Please state 1 if it is never justifiable, 10 if it is always justifiable or a value in between." We also asked subjects to comment on how they filled the income and expenses fields in a text box. We considered four categories when classifying responses: Rule Following; Honesty; Strategic/Evader; Other. While the overwhelming majority of responses fit only one category, some responses fit two categories - often Rule Following and Honesty. The majority of responses (57%) were classified as Rule Following, while the second most coded category was Honesty (27%); 21% of responses were classified as Strategic/Evader and 13% were coded as Other. The proportion of each of the four response categories was roughly constant in all treatments - this means we cannot use those variables to explain treatment level differences in compliance. However, these variables may still be useful to explain choices at the individual level.

Table 1.3 reports the results of a series of Logit models estimating the probability of being a Compliant type. Model 1 considers only the relevant treatments as dummy variables; the omitted treatment is BASE. The basic findings reported earlier are confirmed: the average probability of being compliant is higher in CORR, though the difference is not significant. The reverse is true of OVER (note that our dependent variable equals zero if a subject either under-reports or over-reports tax liabilities), although again, the difference is not statistically significant. The average likelihood of being Compliant is significantly lower in UNDER; that likelihood drops significantly further in the UNDERGENERIC ($\chi^2(1) = 4.50$, p = 0.034) treatment; the likelihood of Compliant in UNDERALWAYS is not statistically significantly different to that in UNDER ($\chi^2(1) = 0.00$, p = 0.997). The likelihood of being Compliant in the UNDERTRIGGER treatment is not significantly different than in UNDER ($\chi^2(1) = 1.89$, p = 0.170), but also not significantly different than BASE.

DV: Compliant	Model 1	Model 2	Model 3
Constant	0.731 **	0.409	0.509
	(0.238)	(0.512)	(0.736)
CORR	0.250	.128	0.388
	(0.338)	(0.348)	(0.435)
OVER	-1.197 **	-1.201 **	-1.468 **
	(0.328)	(0.339)	(0.414)
UNDER	-0.877 **	-0.942 **	-0.864 *
	(0.326)	(0.336)	(0.414)
UNDERGENERIC	-1.584 **	-1.669 **	-1.731 **
	(0.345)	(0.358)	(0.432)
UNDERALWAYS	-0.876 *	-0.982 **	-0.679
	(0.339)	(0.357)	(0.442)
UNDERTRIGGER	0.435	-0.526	-0.315
	(0.334)	(0.344)	(0.423)
Male		-0.232	-0.145
		(0.188)	(0.229)
SelfEmpl		0.258	0.060
		(0.208)	(0.250)
Income		0.010	0.040
		(0.049)	(0.060)
Age		0.009	0.00003
		(0.008)	(0.010)
TaxAtt			-0.118 *
			(0.054)
RuleFollower			1.018 **
			(0.390)
Honest			1.407 **
			(0.382)
Evader/Strategic			-2.242 **
			(0.359)
Other			-1.328 **
			(0.482)
N	554	548	543
Pseudo R2	0.06	0.07	0.31
LL	-359.7	-354.2	-260.0

Table 1.3: Logit estimates of the determinants of Compliant types

Model 2 incorporates participants' observable characteristics, in particular, age, gender through a male dummy variable, a dummy variable for self-employment (SelfEmpl), self-reported annual income (Income) and age. None of these variables are significant, and the sign and significance level of the treatment dummies are unchanged. Model 3 incorporates a variable measuring individuals' attitudes towards paying taxes (TaxAtt), as well as coded responses to the open-ended questions about how they approached their filing decision (RuleFollower; Honest;

Evader/Strategic; Other). Introducing these variables does not change the sign or significance of the coefficients on the other regressors, except for the case of the coefficient on UNDER, which is now significant at the 5% level (p = .037).

The coefficient on TaxAtt is negative and significant: subjects who feel strongly that evading tax is justifiable are more likely not to report their true tax liability in the experiment. In terms of the free-form response coded responses, subjects who described their behavior in terms of following rules or instructions, or who described their actions as a function of honesty were more likely to be Compliant types. In contrast, those who described their actions as evading or strategically grounded were less likely to comply. Finally, those who did not have a clear description of their actions, or who reported not having a clear-cut strategy during the experiment were less likely to be a Compliant type. We conjecture that the last category is capturing some of the individuals who were Under-compliant through error, as opposed to premeditated evasion.

Result 7 *Attitudes towards tax are strongly correlated with under-compliance.*

1.4.5 Revenue

We conclude our analysis by looking at the revenue consequences of defaults and nudges in terms of the level of reported tax liability. Our current analysis has centered on the proportion of individuals who under-report, over-report or correctly report their tax liabilities. It is possible that defaults and other types of nudges we consider in our design, while not significantly changing the proportion of Under-Compliant subjects, will change the amount they under-report[3]. We report on the level of liability declared by subjects, that is the factor that determines revenue before the application of any audit and payment of unpaid taxes or fines, as a proxy for revenue as the figure of actual revenue may be biased by the outcomes of particular audits.

[3]We exclude from this analysis six individuals, who either declared a negative tax liability or a tax liability that meant they would make negative payoffs. In the previous analysis, these individuals would have been classified as Under-Compliant or Over-Compliant; because they are so few, including them did not affect our analysis; however, since tax revenues are continuous, they constitute true outliers, and including them would skew means and standard errors.

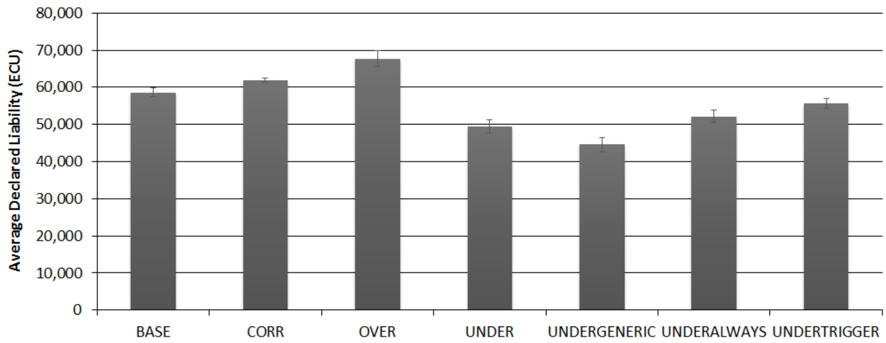


Figure 1.4: Average declared liability

Figure 1.4 displays the average reported tax liabilities in each of the seven treatments. The average tax liability in the BASE treatment was 58,629 ECU. Liability is significantly higher in the CORR treatment (61,936 ECU; MW z = 2.051, p = 0.040). The average liability in OVER is 67,711 ECU and also significantly higher than BASE (MW: z = 4.620, $p < 0.001$), driven by the high proportion of individuals who retained the incorrectly pre-populated entry in their tax form. The average liability in UNDER is equal to 49,456 ECU, significantly less than in BASE (MW z = 3.257, p = 0.001).

The introduction of nudges in the context of UNDER led to a mixed result with regards to liability. The introduction of a checkbox in UNDERGENERIC led to a further reduction in revenues (44,547 ECU; MW: z = 1.985, p = 0.047), while the static descriptive norm message (UNDERALWAYS) resulted in larger average liabilities (52,190 ECU), although the difference relative to UNDER is not significant (MW: z = 0.456, p = 0.648). However, the reactive nudge (UNDERTRIGGER) led to significantly larger average revenues (55,660 ECU) relative to UNDER (MW: z = 2.270, p = 0.023), to the extent that there were no significant differences relative to BASE (MW: z = 1.312, p = 0.189). This suggests that the reactive norm message, while not being able to affect the number of Under-Compliant subjects, did effectively reduce the amount of evasion they engaged in.

Result 8 *Reactive nudges reduced the foregone revenue from evasion from incorrect pre-population of tax returns.*

In short, our data provide only partial support to Hypotheses 4 and 5. While nudges with descriptive norms were generally ineffective at changing the proportion of Under-Compliant individuals, reactive norms did reduce the extent to which Under-Compliant individuals evaded.

1.5 Discussion

Our study seeks to understand how defaults and nudges containing descriptive norm messages shape tax-paying behavior. As such, one of the primary contributions of this paper is to understand the behavioral drivers of honesty in the tax context. In this sense, our work builds on Mazar and Hawkins (2015), who study the extent to which individuals engage in deceptive behavior when it is financially beneficial. That study examines how lying changes as a function of whether the deceptive action was pre-set or not by the experimenter. Mazar and Hawkins (2015) find that deceptive behavior is indeed more prevalent when it is an act of omission. The authors propose that rejecting a correct default is psychologically difficult: it not only involves stating something not true, but also rejecting a pre-existing truthful statement.

Our analysis offers limited support to Mazar and Hawkins' interpretation based on self-concept maintenance. The proportion of compliant subjects in the treatments with correct default values is the same as in our baseline treatment. The time Under-Compliant subjects spent completing the form when the pre-populated value was correct was no different than that spent by Under-Compliant subjects in the baseline condition. However, collected tax revenues in the correct prepopulation treatment were significantly higher than in the baseline treatment. This is only possible if those evading do so by a smaller extent. Importantly, this primarily took place in the non-prepopulated fields. Nudges that displayed a normative message about compliance in response to inputs by subjects raised average declared tax liabilities close to baseline levels.

We provide evidence of different motivations for why and when defaults are hard to override, especially when they are incorrect. For a small subset of individuals in our sample, defaults reduce the cognitive cost of engagement with the filing task. Those individuals will therefore be compliant if the default is correct, but will be under-compliant or over-compliant if the default is incorrect. The behavior of these individuals is therefore consistent with the heuristics approach in decision-making, and supports the claim by Kotakorpi and Laamanen (2015) that pre-populated tax forms reduce the cognitive costs of tax filing.

A large proportion of individuals, when faced with an incorrect default (either under-estimating or over-estimating tax liabilities), responded by evading even more. Subjects may have interpreted incorrect pre-population as incompetence, which could signal greater opportunities for evasion, which is a key determinant of non-compliance in the field (Kleven et al. 2012)

Defaults are extremely powerful: their effect dominates the power of normative messages, which have been shown to be particularly effective in other policy contexts (Cialdini et al. 2006). Only a normative message that was responsive to actual behavior was able to mitigate the adverse effect of an incorrect default, thus providing strong support to the proposal for using reactive defaults made by Smith et al. (2013). Even then, the effectiveness of reactive nudges was primarily on the extent to which subjects evaded, rather than on the proportion of evaders.

One important dimension which we could not explore in the present study is the role of taxpayer trust in the tax agency, and government in general. It is plausible that individuals have low self-efficacy in the tax domain, and choose to trust the values presented in the tax form. This means distinguishing between inertia and self-efficacy explanations of behavior difficult. Self-efficacy could also be at the heart of our revenue result: although the number of under-compliant individuals was not reduced, the amount of evasion was reduced. Future work could look at if the reduction in evasion occurs because those who would have normally accepted the defaults actively change information, or because those who were actively under-compliant actually become more conservative in their evasion.

One important avenue for future research is to further understand how descriptive and injunctive norm information interact with defaults and other forms of nudges in an honesty and/or tax context. Norms and norm abiding are an integral part of one's identity, whether in a social or individual sense. Therefore, by manipulating the extent to which people violate their sense of group identity by evading, we could construct choice environments that deter dishonest behavior.

Finally, our data consist of decisions by UK resident taxpayers. It is possible that attitudes towards government and taxes may differ across different countries. As such, future should investigate the extent to which our results hold in different cultural contexts.

1.5.1 Policy Implications

The experiment detailed in this paper reflects potential differences in the design of on-line tax forms in the UK today and those that may be used in the near future under recent proposals for change. The different treatments reflect situations that might arise under the new filing system relating to the nature and quality of third party reported data used to pre-populate tax forms.

When reviewing the results on an experiment designed to test the effects of defaults on honesty, Mazar and Hawkins (2015) suggest that "it might be even more effective to [...] have tax software automatically pre-fill key fields with available information and require applicants to actively override them rather than typing amounts into blank fields." We argue that this should only be the case for fields for which the tax authority has extremely reliable information. One such case could be employment income: in the UK case, there is already a well-developed system of Pay-As-You-Earn tax reporting, which could be used to pre-populate tax returns of individuals who have multiple sources of income. Bloomquist et al. (2012) report that the quality of such third party data in the US is high, unlike data capital earnings. Our evidence suggests that tax return entries for which government tax agencies have low quality data should not be pre-populated, as this could lead to increases in non-compliance.

We considered measures that the tax authority can mitigate the potentially increased levels of non-compliance, but they should be carefully considered. A lock on the pre-populated field with a nudge for honesty actually caused compliance to worsen if the pre-populated value was below the true level of income. A static nudge containing a descriptive norm message on compliance that was always present had no discernible effect on compliance or revenue.

A reactive nudge reminding users that a lower declaration of income lead to a higher probability of audit was much more effective in increasing compliance, particularly in relation to the major income item in subjects' profiles. We note that the message we used in this experiment is highly specific to the profile used. Generating an equivalent message in a real tax system would be non-trivial for a tax authority.

A noteworthy effect of the reactive nudge is that it increases response times by Compliant types, while having a limited effect on those under-reporting their tax liabilities, both in terms of reducing their number or changing their deliberation time. This introduces a welfare question: in a more complex environment where the filing task could be expected to take days rather than minutes, this nudge is introducing an extra burden on compliant types, while not necessarily reducing the number and behavior of non-compliant types.

More generally, our findings indicate an important scope for nudges that react to users' behavior. This types of nudges are particularly well suited to online environments, where we expect most of the interactions between people and their government or private service will take place in the near future. The fact that nudges react to user behavior may lead users perceiving that the nudge is directed at them, thus further increasing their potential. This is a promising area for future policy implementation.

Chapter 2

An Experimental Test of a Simple Psychological Model of Lying Aversion

2.1 Introduction

Every day people are faced with the temptation to tell lies. Occasionally such opportunities may be for personal gain, but more typically they are directed towards the upkeep of our own self-image or that of others. Indeed the majority of lies people reported telling in diaries kept to track dishonest behaviour were for psychological reasons such as to appear kinder or smarter than they were or to avoid embarrassment or conflict (DePaulo et al. 1996). Cultures may well contain norms of politeness that require the ability to tell some forms of lie; for example the ability to show thanks for an unwanted gift is often encouraged and developed in childhood (Talwar et al. 2007). In the same manner, the art of impression management, the act of influencing the perceptions of oneself by other people requires different forms of deception, such that people are often well versed in small everyday lies (Goffman and Best 2005). Indeed such lies are typically easily told, arousing little discomfort in the teller (DePaulo et al. 1996).

Serious lies are, however, typically considered somewhat differently, often creating serious distress for the person telling the lie as they must break a perceived social norm of honesty (DePaulo et al. 1996). There are, however, numerous examples of serious lies having a large economic impact, such as incorrect accounting information being used to inflate values by companies such as Enron, Autonomy and Tesco (Financial Times 2012, 2014; Healy and Palepu 2003) and misreported facts designed to attract custom or investment by firms including Facebook, Snapchat and VW (Financial Times 2017; New York Times 2014; The Guardian 2017). The sum of small or moderate lies can also add up to form significant measures. The sum of all criminal attacks, deliberate evasion and activity in the hidden economy designed to avoid or reduce tax payments amounted to contribute £15.6 billion to the tax gap of missing liabilities in the UK for the year 2013-2014 (HMRC 2015b). As well as producing a direct monetary cost, lying can lead to the indirect effect of wearing down the trust on which advanced economies are built. It has been suggested that the existence of a social norm of trust, including the need for truthful disclosure, may serve as a mechanism to allow beneficial cooperation that would otherwise be too expensive to achieve (Arrow 1970).

For many people the act of lying therefore constitutes a social dilemma. There are occasions when lying is acceptable, even encouraged, and others when lying is seriously frowned upon, so people often learn to be able to choose when and how to lie. Consider for example a job applicant at an interview. Among the questions asked, they may be asked for their IQ. This question is different in nature to one relating to the applicant's academic record as it is typical in a recruitment process to have to produce certificates to prove an academic record, whereas it is not typical to validate an IQ score. If the applicant knows their score, they have the potential to report a higher value in a bid to appear smarter. The decision that the potential employee then faces is how much should they increase their IQ score by? For a person with a below average IQ score, there is a temptation to at least increase the score to the average value, if not slightly above, to at least match the perceived average of the competition for the job. There is no such inventive for those with an IQ score above the average, though there still remains some incentive to report a value higher than the poorer, or even equivalent, competition. However, there is a potentially a limit to the IQ score that an applicant should report, in part owing to the believability of any value reported, but also as they may not wish to appear dishonest. Furthermore an applicant may limit the value they report in a bid not to appear boastful. For a job applicant to boost their IQ score a little bit in response to the question seems innocuous, and indeed a rational response if an applicant believes that others will do so, but to what degree would an individual lie and does the extent of misreporting differ with their actual true score?

The example of giving thanks for an unwelcome gift at the beginning of the chapter indicates that people are often conditioned on when to lie from an early age. The key element of the model presented in this chapter addresses a potentially alternative dimension of such conditioning, that of how to lie.

Studies in experimental economics have found that subjects in the laboratory have an aversion to lying as participants select less self-serving outcomes when it requires the use of a deceptive message than through a direct choice (Gneezy 2005). Subjects have also been found to over-report their outcomes from unobserved random events, such as a dice roll or a coin toss, when compared to the theoretical outcome, but not always to the maximal degree (Cohn et al. 2014; Fischbacher and Föllmi-Heusi 2013). Experimental subjects were also found to report higher values for a calculation solving task when self-reporting their scores compared to where the results were checked by the experimenter, but very few subjects cheated by reporting the maximum amount (Mazar et al. 2008).

Recent research has begun to focus on a model for lying aversion, with a particular emphasis of the effect of reputation on the agent's decision (Abeler et al. 2016; Gneezy et al. 2016). Abeler et al. (2016), compare a variety of possible models for lying aversion and perform experimental tests of the predictions from the models. The authors report that only a model combining a preference for being honest with a preference for being seen as honest can explain their data.

This finding is in keeping with separate experimental results detailing a preference for being seen to be honest presented by Hao and Houser (2017).

The observation of a model relating to a preference for appearing honest relates to the expected intuition of the motivating example of a job applicant reporting their IQ. The reputation element in the model serves to create a value to report that reflects the subject's preference for being seen to be honest. The approach in this chapter differs from that of Abeler et al. (2016) in two main ways.

The first main difference relates to the model, specifically the nature of the cost function used to express lying aversion in the model. As expressed in the lying cost - reputation for honesty model of Abeler et al. (2016), the model presented here incorporates two terms to express the cost of lying, the first of which is also a cost in terms of the size of the lie. The model differs in the second cost term which is framed in terms of some deviation of the value reported by the individual from a reference point. The reference point can be considered as some value known or estimated by the agent to represent the average of the value or as a measure of the values reported by others. The reference point can then be viewed as an anchor for the agent's lying cost in terms of a preference for appearing honest.

Alternatively, the reference point can be interpreted as the point above which reporting a value becomes less believable. An individual facing a decision to falsely report some value in a manner that can increase their utility may believe that reporting a value well above the average value is less credible than simply reporting a value slightly above the average. Returning to the example, when questioned as to their IQ, stating 120 for a job applicant of average IQ (with a true score of 100) may boost their employment chances without comment, whereas giving a value in excess of 130 may arouse some degree of suspicion in the potential employer. Alternatively, the reference point may represent a point above which the reported value may appear to be boastful and the act of bragging may cause disutilty. For example, it may be that the potential employee has a very high IQ, but does not wish to broadcast that fact.

The model presented in this chapter therefore contains two psychological cost terms reflecting different sources os disutility. The first captures any degree of intrinsic dislike for lying by an individual. It may simply be that an individual dislikes lying, and, as is assumed here, that dislike increases with the extent to which the individual deviates from the truth when making a false report. Such a dislike for lying, however, need not necessarily exist for all individuals, indeed some may experience little or no cost from lying through this mechanism[1]. The second lying cost term captures any dislike for reporting an unbelievable lie, and further assumes that such a dislike is increasing in the value reported as the lie become less believable. The reference point captures the point at which the value reported starts to loose believability. As with the first cost term, it may be that any particular individual may experience little or no such a cost in terms of deviation from the reference point. The model therefore allows for any combination (neither, both or just one) of the two lying cost terms to determine an individual's decision.

The second main difference from the approach of Abeler et al. (2016) presented in this chapter is the use an alternative test of the model. As in Abeler et al. (2016), experimental subjects were required to undertake some task that generated a score that they were required to pay themselves for, leading to an opportunity to cheat through an implied misreporting of the value. There are three elements to the difference in the experimental approach described in this chapter. Firstly, a more natural setting was used, whereby subjects undertook a numeracy test to produce scores, rather than a have a lottery generate the values to be reported. This design provides a less abstract setting for the lying decision and allows for some element of error, in that subjects can incorrectly answer questions on the test and still claim an associated payment. Secondly, the design allowed for an assessment of cheating at the individual level as the test papers of the subjects to be collected, while actions were not attributable to any particular individual as payments were kept anonymous. Thirdly, as the numeric ability of the subjects was unknown, the experimental design uses a between subjects comparison of two treatments with different reference points based on the modal

[1] It is also possible that some individuals experience extra pleasure from telling lies, such individuals are classed as having a zero cost in the discussion here

scores of a prior control treatment to serve as a variation in the reference point, rather than using alternative distributions of an underlying lottery.

2.2 Literature

There is considerable experimental evidence that some people are averse to lying. Gneezy (2005) observed that higher proportions of people choose a self-serving outcome from a pair of payoffs when presented with the choice in a dictator game compared to in a sender-receiver game where obtaining the greater payment for oneself required telling a lie. A similar result is presented by Cai and Wang (2006), who report over-communication of the truth in sender-receiver games compared to a sequential equilibrium model. Sánchez-Páges and Vorsatz (2009) present results from a sender-receiver game with a silent option and conclude that the over-communication of truth in such games arises from an aversion to lying rather than a preference for truthfulness.

The results of Gneezy (2005) may arise due to a population of two distinct types, a first of people who are morally driven and never lie and another of individuals who are economically driven and will lie to obtain a preferred outcome, or from a population with heterogeneous costs of lying. Hurkens and Kartik (2009) report an extension of the experiment of Gneezy (2005) and present evidence from their results that fails to reject the hypothesis of two types. Gibson et al. (2013) present evidence for heterogeneous preferences for truthfulness from an experiment where subjects face the same reporting decision but with different levels of economic cost. The authors conclude that the variation in the choices made by subjects at different levels of cost is consistent with heterogeneous lying costs.

The variation in peoples' choices with the type of lie has been examined by Erat and Gneezy (2012), who report different proportions of subjects choosing to tell a lie in a sender-receiver game when the payoffs associated with the lie vary as to which of the players may benefit from the lie. A higher proportion of subjects were recorded telling a lie that benefited both players, though 35% of subjects in their still did not tell such a Pareto-improving lie. Vanberg (2017) presents evidence that the extent of the lying aversion observed in this condition may be due to sender's expectations of the receiver's actions in the experimental design.

The role of a subject's beliefs about the beliefs of the other player in a sender-receiver game have also been examined. López-Pérez and Spiegelman (2013) conclude that subjects' behaviour is more consistent with pure lie aversion than a model of belief dependent lie aversion. Peeters et al. (2015) provide evidence that over-communication of the truth correlates to the sender's first order beliefs about what a receiver will do, suggesting some subjects exhibit a preference for truth telling. The authors find no evidence that the sender's action correlates with their second order beliefs, suggesting excessive truth telling is unlikely to be driven by guilt aversion in their experiment.

Other studies have examined the role of the effect of various contexts on the decision to lie, for example of ex-post disclosure (Behnk et al. 2014; Greenberg et al. 2015), of a competitive or co-operative context (Rode 2010), of social ties (Chakravarty et al. 2011), of the communication channel (Conrads 2014; Holm and Kawagoe 2010), of framing of feelings (Cappelen et al. 2013) and of correlation with pro-social behaviour (Biziou-van-Pol et al. 2015).

The papers discussed thus far have examined lying under the condition where the experimenter knows the true state, largely in the setting of a sender-receiver game. An alternative strand of the literature examines lying under the condition that the experimenter does not know the true state, and while they are therefore unable to examine the individual's decision to lie, they can examine lying at the sample level by comparison to an underlying theoretical distribution or a measured benchmark.

Fischbacher and Föllmi-Heusi (2013) present results from an unobserved dice rolling task appended to other experimental sessions in which subjects were offered different payments for the values they could report. Subjects were found to over-report both the highest payoff and the second highest payoff scores compared to the theoretically expected level. The authors term the over-reporting of the second highest payoff score as partial lies, in that subjects lied, but did not report the highest score possible. In a similar study, Utikal and Fischbacher (2013) found that nuns under-reported their scores from a dice roll compared to female students.

Mazar et al. (2008) offer evidence in support of a theory of self-concept maintenance, in which people are able to tell lies under conditions where they do not have to negatively update their self-image. The authors detail results from a number of experiments based on a puzzle task whereby subjects must find matching pairs of numbers within a matrix, where cheating is possible in treatments that are self marked by subjects. An increased attention to standards, in the form of a task remembering the ten commandments or a reminder of the honour code of the student's institution, was found to reduce the level of cheating compared to neutral treatments.

One result of direct interest to the work presented here arises in the fifth experiment of the paper. The authors present a two-by-two treatment where in the first dimension the ability to cheat was varied (control versus recycle) and in the second dimension subjects were told different values of average number of calculations solved by subjects on the test, in one treatment they were told the correct value of four while in another they were told an exaggerated value of eight. The results show a significant effect for the difference in the ability to cheat, but no effect arising from the change in reported average performance.

Abeler et al. (2014) describe an experiment where members of the German public were contacted by telephone by a market agency to undertake a survey. A potential payment for the survey was offered at the end with the outcome to be decided by an unobserved coin toss or tosses by the subject. The results show no significant deviation from theoretical predictions, suggestive of a high degree of honesty among the sample contacted. Student subjects in the laboratory were found to lie more, though as in the results of Fischbacher and Föllmi-Heusi (2013), not by only over-reporting the maximum case but also a lesser payoff. Cohn et al. (2014) also report on an experiment with a repeated coin toss and found that subjects who work in the banking industry were more dishonest when their identity as part of banking culture was made salient prior to be asked to perform the coin tosses.

An interesting insight into the process of lying is offered by Shalvi et al. (2011b) who give an account of a dice roll experiment where subjects were told to roll the dice three times and report the value of the first roll. The authors describe how the experimental results are consistent with a model of a best of three rolls, suggesting that the subjects may have used the counterfactual rolls observed in determining their actual reported value.

The papers investigating reports of values unobserved by experimenters highlight a pattern of incomplete lying. Some subjects do lie, but the lies told are not always the maximal ones. Different results have been observed in terms of behaviour, from largely truth telling in the coin flip of Abeler et al. (2014), to consistent small lies in the matrix test of Mazar et al. (2008), which hints that context forms an important part of the decision to lie. A further strand of the literature seeks to examine the effect of justification by allowing subjects to cheat "in the mind", in that they are asked to make some prediction about an event before it happens but only asked to report on that view after it has been observed.

Jiang (2013) reports significantly higher levels and values of cheating in an experiment where subjects were asked to choose the face up or face down side of a die for payment between a treatment where subjects were only asked to write down their choice after the dice roll compared to one where they were asked to write their choice down first. The author concludes that the behaviour required to cheat in the write-first treatment serves as a source of greater hurt to subjects' self-image than in the throw-first treatment. Hao and Houser (2017) report an experiment where lying is significantly reduced in a treatment where cheating is perceived as planned, as the act of cheating requires a pre-meditated step, compared to where the act may be viewed as impulsive. The authors argue that their results are consistent with a view of people having a preference for appearing honest than an intrinsic preference for honesty. While both experiments show that there is more lying when the intent of such an action is less transparent, it is unclear if this result is due to a preference for appearing honest or due to a perception of greater opportunity to cheat.

The experiments presented so far give evidence that people often avoid telling lies and suggest that such lying aversion behaviour relates to people's ability to self-justify their actions. Much of the evidence matches to the premise presented in the introduction that people face many opportunities for dishonesty and have mechanisms for choosing how to behave such that outcomes may vary with the context. For example, the degree to which a dishonest action can be self-rationalised in terms of the relative benefits to oneself and to others has been shown to alter outcomes. Moral reminders can serve to manipulate people's decisions and reduce dishonesty. Differences in the chance of being caught or in the role of the perception of others about the intention of a dishonest action, as to whether pre-meditated or impulsive, can also affect people's choices in the decision to report falsely. A key piece of evidence presented is the observation of partial lies, that is outcomes where people report falsely but do not simply report the payoff maximising option. The final literature section therefore focuses on a set of experiments that examine the extent of lying, that is where subjects have some reward that is increasing in the size of the misreported value.

Lundquist et al. (2009) describe an experiment in which the size of the lie was found to have an impact where subjects were required to report their score on a task to another participant and low scoring subjects had an incentive to over-report. The authors detail that the probability of a lie increased the actual score (for those with the appropriate incentive to lie) concluding that the aversion to lying increases with the size of the lie. Gneezy et al. (2013) report on an experiment based on a sender-receiver game where the payoff to the sender is a function only of the message sent and not of the action of the receiver. The results shows that the propensity of the subjects to report the maximum value falls as the actual value increases, from which the authors conclude that there is considerable aversion to lying and that there result is inconsistent with a theory based on the size of the lie.

Gneezy et al. (2016) examine a model of concave lying costs with a term for reputation using an experiment with a random draw. The authors argue that the fall in the lying rate and a high proportion of maximal lies is consistent with a model of a fixed cost of lying and not with a model of a convex cost of lies. The results also show that reputation matters as more subjects report more partial lies when outcomes are not observed by the experimenter. The authors conclude that there is an important impact of the size of the lie in terms of reputation.

The results of Mazar et al. (2008) are also consistent with a model based on the size of lie, subjects do not report the maximal value, suggesting there is some optimal lie for the value they wish to report that is less than the maximum. Shalvi et al. (2011a) report an experimental result whereby subjects are willing to tell a medium lie but avoid a small or a large lie, suggesting that they find the justification of the medium lie easier.

In a recent paper, Abeler et al. (2016) perform a meta analysis of a large number of experiments with reports of unobserved lotteries. The analysis reveals that the partial lying result of Fischbacher and Föllmi-Heusi (2013) is robust over the large number of experiments surveyed, in particular that subjects report over the whole range of values, and that non-maximal payoff reports are made more often than would be predicted. The authors propose a series of different models and asses them against a set of criteria designed to reflect the main stylized facts of the meta analysis. The models incorporate terms for lying costs, a reputation for honesty and social norms as well as a number of sub-variants for each of these categories and combinations. The authors describe a series of experiments designed to test the predictions of their models and conclude that a model containing a lying cost and a reputation for honest cannot be rejected by their data.

The literature presented in this section details a number of experiments investigating the propensity of subjects to lie. A key result observed across experiments is that of lying aversion, in that subjects are often more truthful than either rationality or their social preferences would predict. A number of the results suggest that reputation has a role to play in people's decision as to whether or not to lie. There are also a number of differences, in particular as to whether or not the size of a lie has an effect of behaviour.

The results also indicate that ancillary features of the experimental design, such as the nature of the information being lied about, the perception of the opportunity to cheat, the believability of any value reported and impact of the discovery of lie, whether financial or non-pecuniary in terms of shame or reputation are all potentially important for behaviour. The results give rise to particular concern with regard to the experimental design. In some settings it may be that once it becomes clear to a subject that cheating is possible, their decision collapses to a binary one, cheat using the maximal value or report the truth. This may be consistent with particular real world situations but, as argued in the introduction, not with all the circumstances in which people have the opportunity to lie. The use of an abstract lottery, such as a random draw by a computer or a dice roll, to produce values may further promote such polarisation of the decision. Such designs may serve to much reduce the roles of ownership and believability in people's decisions - as evidenced by the significant proportion of people reporting they achieved 10 straight winning flips of a coin (Cohn et al. 2014). The results presented, however, also show that ownership and believability can change behaviour and therefore may have an important role in appropriate contexts.

2.3 Theory

2.3.1 Model

The model considers a set of states distributed over a normalised interval from zero to one, $S = [0, 1]$. A state of the world, $s \in S$ is drawn from a differentiable distribution function $F(s)$ and privately revealed to an agent. The agent must then report a value for the state, $r \in S$, that determines their payoff, $v(r)$. There are two channels through which an agent experiences disutility from lying, a cost in terms of the size of the deviation of the reported value from the true value, $c(r-s)$, and a cost relating to the deviation above the reference point, $c^R(r-s^R)$. The reference point, $s^R \in S$, can be viewed as an anchor for the agent's lying cost in terms of a preference for appearing honest, or alternatively the reference point can be interpreted as the point above which a reported value becomes less believable. For simplicity it is assumed that the disutility from reporting a lie is separable from the benefit. The utility function of the agent in the model is therefore expressed by:

$$U(r,s,s^R) = v(r) - c(r-s) - \mathbb{1}c^R(r-s^R) \tag{2.1}$$

where $\mathbb{1}$ is an indicator function that takes the value 1 if $r \geq s^R$ and zero otherwise. The agent faces the problem:

$$\max_{r \in S} U(r,s,s^R) \tag{2.2}$$

The following assumptions will hold through the discussion.

Assumption 1 *The payoff to the agent is continuous and strictly increasing in the value reported*

Assumption 2 *The cost function* $c(r-s)$ *is continuous, zero when there is no lie and is either zero everywhere or is strictly convex with a minimum at the point of truth*

Assumption 3 *The cost function* $c^R(r-s^R)$ *is continuous, zero when the value reported is less than or equal reference point and is either zero everywhere or strictly increasing and convex for reported values above the reference point*

The first assumptions states that the payoff function, $v(r)$, is continuous and increasing in the value reported ($v(r') > v(r) \quad \forall r' > r$) such that reporting a higher value increases utility in the absence of any lying costs. The second assumption states that agents are either unconcerned by any form of lying, or their concern increases with the size of the lie at an increasing rate. That is $c(r - s) = 0$ if $r = s$, and either $c(r - s) = 0 \quad \forall r$ or $c(r' - s) > c(r - s) \quad \forall r' < r \leq s$, $c(r' - s) > c(r - s) \quad \forall s \leq r < r'$ and $\lambda c(r - s) + (1 - \lambda)c(r' - s) > c(\lambda r + (1 - \lambda)r') - s) \quad \forall r' > r, \lambda \in (0, 1)$. The third assumption states that agents are either unconcerned by telling a lie that involves reporting a value above the reference point or face a cost that is increasing in the size of the lie above the reference point at an increasing rate. That is $c^R(r - s^R) = 0$ if $r \leq s^R$ and either $c^R(r - s^R) = 0 \quad \forall r > s^R$ or $c^R(r' - s^R) > c^R(r - s^R) \quad \forall s^R \leq r < r'$ and $\lambda c(r - s) + (1 - \lambda)c(r' - s) > c(\lambda r + (1 - \lambda)r') - s) \quad \forall r' > r > s^R, \lambda \in (0, 1)$. The assumptions in terms of the two costs from lying are similar, in that both allow for an agent who is unconcerned by that particular form of cost and state that the cost for those who do have a concern is increasing with positive deviation. The two lying costs terms differ in that, for any agent with a concern for the particular cost, the cost in terms of the size of a lie applies to misreported values above and below the true value, whereas the cost in terms of deviation from the reference point applies to values reported above the reference point.

The continuity component of the assumptions given above ensure there is a solution to the agent's problem that can be expressed as the correspondence $r^*(s, s^R)$.

2.3.2 Predictions

The first proposition examines the change in value reported by an agent upon an increase in the state.

Proposition 1 *The minimum value reported by an agent is non-decreasing in the realised state*

The minimum value detailed in the proposition refers to the lowest value of the optimal range the agent will report ($r^*(s, s^R)$) for some given pair of a state and a reference point. For two states where one has a larger value than the other, $s_2 > s_1$, and an unchanged reference point, this proposition states that the optimal solutions for the value to report for the two states, $r^*_2(s_2, s^R)$ and $r^*_1(s_1, s^R)$, are such that $\min(r^*_2(s_2, s^R)) \geq \min(r^*_1(s_1, s^R))$. A proof is given in the appendix.

This result ties to the predictions of monotone comparative statics (Milgrom and Shannon 1994). The only component of the utility function affected by a shift in the state is the cost in terms of the size of the lie. The convexity assumption on the cost term $c(r - s)$ leads to decreasing differences, that is as the state is raised the difference in the value of the cost function for any two fixed values to report falls. When the cost term is applied as a negative term in the utility function, this leads to increasing differences, and the monotone comparative statics result that the optimum is non-decreasing can be applied.

The first proposition relies on continuity of the payoff function given in the first assumption but not on the assumption that the function is increasing. The proposition still applies in the case where the payoff is decreasing, that is where under-reporting the true state is beneficial to the agent. The second proposition uses the assumption of increasing payoff to determine lower bounds on the value the agent will report.

Proposition 2 *For states below the reference point, the value reported by an agent must be at least the true state. For states on or above the reference point, the value reported must be at least the reference point*

For states below the reference point ($s < s^R$) reporting a value below the true value can never form a solution as it will always lead to a lower level of utility than reporting the truth. For states on or above the reference point ($s \geq s^R$) reporting a value below the reference point can never form a solution as it will always lead to a lower level of utility than reporting the reference point. Reporting a value below the true state may, however, lead to a higher utility than reporting the truth as there the reduction in the cost in terms of deviation from the reference point may be greater than the additional cost of in terms of reporting a lie and the loss of payoff from reporting a lower value.

This proposition details an important consequence of the inclusion of the reference point. An agent who is given a true state below the reference point will never report a value that is less than the true state. However an agent that receives a state above the reference point may report a value that is below the true state but that must still be on or above the reference point.

The first two propositions can be combined to characterise the lower bound of agent behaviour for a given reference point. For a given state, the minimum value that an agent might report will be given by the either the true value of the state or the reference point depending on whether the state is below or above the reference point. For any increase in the state, the lower bound of the value reported by the agent will be at least the lower bound of the values potentially reported by the agent at the previous state.

Proposition 3 *The minimum value reported by an agent is non-decreasing in the reference point*

For two reference states where $s_2^R > s_1^R$, the lower bound of the optimal solutions are given by $r_1 = \min(r_1^*(s, s_1^R))$ and $r_2 = \min(r_2^*(s, s_2^R))$. The proposition states that $r_2 \geq r_1$. A proof is given in the appendix, derived in the same manner as for the first proposition. This proposition states the cost of lying associated with some reported value for a given realised state will not rise under an increase in the reference point, such that the agent will either report the same value or potentially a higher one. An increase in the reference point may serve to make a misreported value above the reference point a more credible lie or to appear less boastful.

2.3.3 Model Extension

The analysis presented in the previous section makes clear predictions with a limited number of assumptions. The propositions clearly define the lower bound of behaviour, in that the minimum values reported by an agent can be determined and the change in the lower bound of the value report with a change in the state or a change in the reference point can also be characterised. It is, however, difficult to make further concrete predictions, in particular relating to any potential upper bound on behaviour in relation to the change in state, without additional assumptions. In the ensuing discussion, the following additional assumptions will be used.

Assumption 4 *The payoff function is twice differentiable and weakly concave in the value reported*

Assumption 5 *The cost function in terms of the size of the lie is twice differentiable other than at the point of reporting the truth*

Assumption 6 *The cost function in terms of the deviation from the reference is twice differentiable other than at the point of reporting the reference point*

The first of the additional assumptions can be stated as $v'(r) > 0$ and $v''(r) \leq 0$. This assumption allows for constant or diminishing marginal utility of the payoff. The second additional assumption, stated in terms of the size of the lie $x = r - s$, can be written as $c^{'}(x) < 0$ for $x < 0$ and $c^{'}(x) > 0$ for $x > 0$ and $c^{''}(x) > 0 \quad \forall x \neq 0$ other than the case where $c(x) = 0$. The discontinuity of the derivative at the point of no lie ($x = 0$) allows for truth telling behaviour. The third additional assumption can be expressed as $c^{'R}(r - s^R) > 0$ for $r > s^R$ and $c^{''R}(r - s^R) > 0$ other than the case where $c^R(r - s^R) = 0$. There may be a discontinuity in the first derivative at the reference point in that it need not be zero, allowing for behaviour whereby agents report the reference point.

2.3.4 Predictions for Model Extension

It can be noted that under the additional assumptions the utility function, other than in the case where the two cost terms are zero everywhere, is strictly concave. As the case where the costs are zero everywhere leads to the trivial solution $r^*(s, s^R) = 1$ for all s and s^R, then under the additional assumptions there is a single optimal value for the agent to report for each value of s for a given value of s^R. The propositions of the previous section can therefore be restated in terms of the value reported by the agent, rather than the minimum value of the range of values reported.

The solution to the agent's problem may be characterised by the appropriate first order derivative condition, though there are three important boundary conditions to consider, in particular relating to payoff maximising behaviour and to behaviour relating to the discontinuities in the derivatives of the cost terms.

The second proposition of the previous section detailing the lower bound on the value an agent will report can be recast in terms of the marginal benefit and marginal cost. For realised states below the reference point, $s < s^R$, the marginal cost of reporting a state below the true value is negative and as the marginal benefit is assumed to be positive reporting values below the reference point cannot form a solution to the problem. Thus $r^*(s, s^R) \geq s$ for $s < s^R$. For realised states above the reference point, it may be that the boundary condition for reporting the reference point, $v'(r) \leq c'^R(r - s^R) \forall r \geq s^R$, applies, such that $r^*(s, s^R) \geq s^R$ for $s \geq s^R$. The next proposition uses the additional assumptions to examine an upper bound on the agent's behaviour upon a change in the state.

Proposition 4 *The maximum increase in the value reported by an agent upon an increase in the state is equal to change of the value of the state*

A proof of this proposition can be derived from the first order condition for an interior solution to the agent's maximisation problem. This can be expressed by the following:

$$v'(r^*(s, s^R)) - c'(r^*(s, s^R) - s) - c'^R(r^*(s, s^R) - s^R) = 0 \tag{2.3}$$

Denoting the derivative of the optimal value r with respect to the state as r', taking the second derivative yields:

$$v''(r^*(s, s^R))r' = c''(r^*(s, s^R) - s)(r' - 1) + c''^R(r^*(s, s^R) - s^R)r' \tag{2.4}$$

By assumption c'' and c''^R are both non-negative whereas v'' is non-positive, so it must be that $r' \leq 1$.

This proposition gives the upper bound on how the value reported by the agent may change with an increase in state in that it can be at most the same as the change in the value of the state. Furthermore, proposition one can be expressed as $r' \geq 0$, such that the range of values that the value reported by the agent with a change in the state can be expressed by $r' \in [0, 1]$. It can be noted that the three boundary conditions for which the first order condition does not hold conform to this range for a change in the reported value with a change in the state. In a first boundary condition, an agent may continue to report the maximum value, whereby $r' = 0$. In a second, an agent may continue to report the truth, such that $r' = 1$. In the third boundary condition an agent may continue to report the reference point, such that $r' = 0$.

An interesting feature to note is that where $c''^R(r^*(s, s^R) - s^R) = 0$, such as would apply when $r < s^R$, that is the agents optimal value would be below the reference point, then the case where $r' = 1$ corresponds to the condition $v'' = 0$. That is the agent will report a value that has a constant difference to the state where they have a constant marginal utility with regard to the payoff for reported values below the reference point.

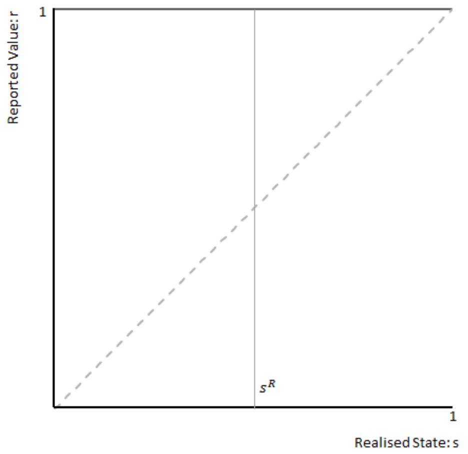
2.4 Types

There are a number of different types of agent that exhibit varying forms of behaviour that are consistent with the model. This section will describe a number of them. The first six types are illustrated in Figure 2.1. These six types represent the more extreme cases consistent with the model.

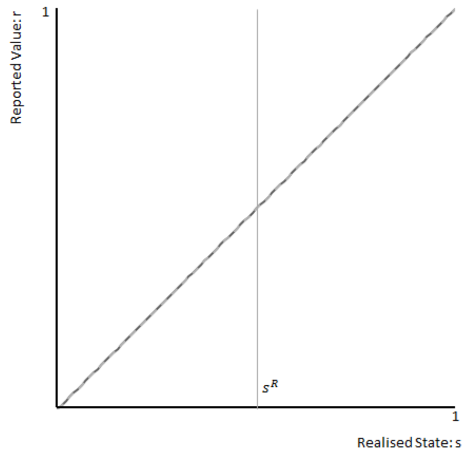
Behavioural Type 1 *A dishonest type of agent may exist that always reports the maximum value,* $r^* = 1$

The first type reflects a rational agent with little or no cost of lying such that they always report the maximum value. The dishonest type of agent is illustrated in Figure 2.1a. The following condition characterises such a type of agent:

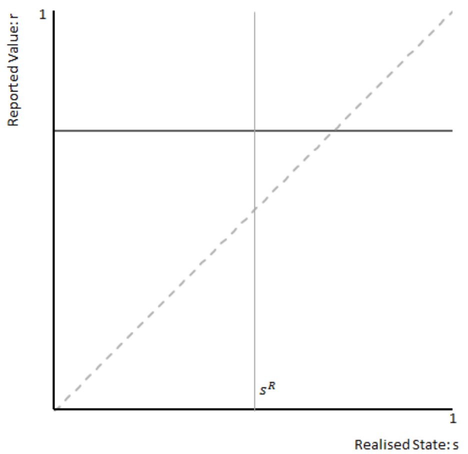
$$U(1, s, s^R) > U(r, s, s^R) \quad \forall r \in [0, 1); \quad \forall s \in S \tag{2.5}$$



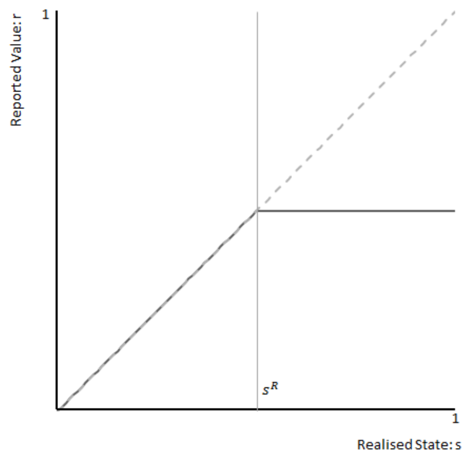
(a) Dishonest



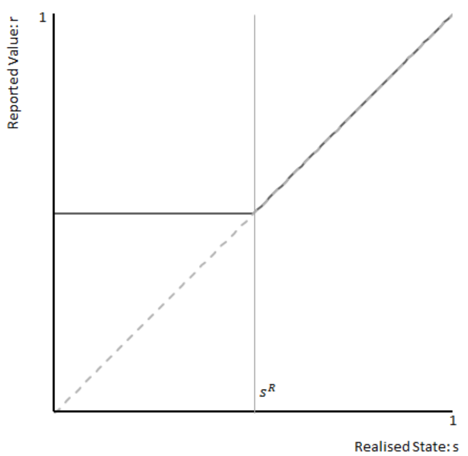
(b) Honest



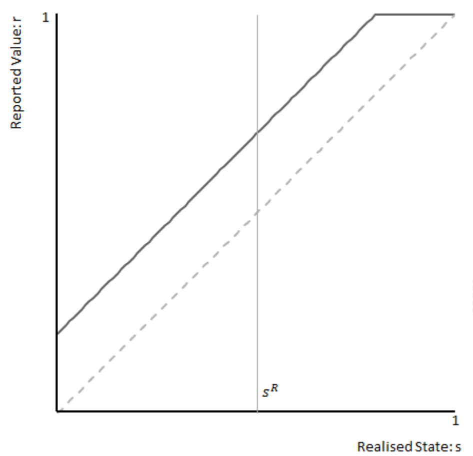
(c) Fixed Value



(d) Honest Modest



(e) Dishonest Modest



(f) Fixed Lie

Figure 2.1: Illustration of behavioural types

Behavioural Type 2 *An honest type of agent may exist that always reports the true value,* $r^* = s$

The honest type has a sufficiently high cost of lying in terms of the size of the lie that they always report the true value. The honest agent is illustrated in Figure 2.1b. The following condition characterises such a type of agent:

$$U(s, s, s^R) > U(r, s, s^R) \quad \forall r \in [0, 1] \quad \& \quad r \neq s; \quad \forall s \in S \tag{2.6}$$

Behavioural Type 3 *A fixed value type of agent may exist that reports the same value for all states*

The fixed type of agent experiences a cost in terms of deviation from the reference point such that they will always report the same value. The fixed value type of agent is illustrated in Figure 2.1c and is characterised by the following condition:

$$U(r^T, s, s^R) > U(r, s, s^R) \quad \forall r \in [0, 1] \quad \& \quad r \neq r^T; \quad \forall s \in S \tag{2.7}$$

where r^T is the fixed value that the agent reports. The value r^T will be discussed further in the next section.

Behavioural Type 4 *An honest modest type of agent may exist that reports the true value for states below the reference point, but the reference point for states on or above the reference point*

The honest modest agent is illustrated in Figure 2.1d and is characterised by the following conditions:

$$U(s, s, s^R) > U(r, s, s^R) \quad \forall r \in [0, 1] \quad \& \quad r \neq s; \quad \forall s \in [0, s^R) \tag{2.8}$$

$$U(s^R, s, s^R) > U(r, s, s^R) \quad \forall r \in [0, 1] \quad \& \quad r \neq s^R; \quad \forall s \in [s^R, 1] \tag{2.9}$$

An honest modest agent can be thought of as being generally driven by honesty, such that they report the truth in low states, but whose desire to conform to reporting the reference point is sufficiently strong that it outweighs their desire for honesty and causes them to misreport their value by under-reporting the true value for states above the reference point. The honest modest agent wishes to be honest but even more wishes to *appear* honest for high realised states.

Behavioural Type 5 *A dishonest modest type of agent may exist that reports the reference point for states below the reference point, but the truth for states on or above the reference point*

The dishonest modest type is a striking alternative to the honest modest type given before. The dishonest modest type is illustrated in Figure 2.1e and is characterised by the following conditions:

$$U(s^R, s, s^R) > U(r, s, s^R) \quad \forall r \in [0,1] \quad \& \quad r \neq s^R; \quad \forall s \in [0, s^R) \tag{2.10}$$

$$U(s, s, s^R) > U(r, s, s^R) \quad \forall r \in [0,1] \quad \& \quad r \neq s; \quad \forall s \in [s^R, 1] \tag{2.11}$$

The dishonest modest agent experiences the cost of lying from both cost components of the utility function, but the cost in terms of the reference point is dominant, such that for low states the agent is prepared to report a lie, but only for states up to the reference point. For states above the reference point, the marginal cost for states between the reference point and the true state is reduced to a level below the marginal benefit by the contribution from reporting a value that is below the true state in terms of the size of the lie such that reporting the true state becomes optimal for the agent. A dishonest modest agent does not mind telling a lie, but finds it beneficial to report the truth for states above the reference point.

Behavioural Type 6 *A fixed lie type of agent may exist that reports a lie of a constant size for states that they are able to do so and otherwise will report the maximum value*

The fixed lie type of agent is illustrated in Figure 2.1f. Such an agent has no concern for the reference point (such that $c^R(r - s^R) = 0$ everywhere) but some cost in terms of the size of the lie such that for some values of the state the agent's choice is characterised by the interior optimum. Furthermore the agent has a constant marginal benefit from the payoff (v), such that the solution to the agent's optimisation problem is a fixed value, r^{+}, with respect to the state so long as the implied value to report is less than the maximum value. This agent is therefore characterised by the conditions:

$$v'(r^{*}) = c^{'}(r^{*} - s) \quad v'(r) = v, \quad \forall s < 1 - r^{+} \tag{2.12}$$

$$r^{*} = 1 \quad \forall s \geq 1 - r^{+} \tag{2.13}$$

The condition given for a fixed lie type of agent always has the same solution relative to the point of truth (so long as there remains an interior solution), such that the agent will therefore be comfortable with a particular size of lie to tell. A fixed lie type will embellish the realised value to the same extent whatever the realised value up to the point where they are reporting the maximum value anyway. The main proposition gives that the reported value is non-decreasing in the realised state, such that once an agent reports the maximum value, they will report that value for all higher states.

The fixed lie type reveals a further important result of the model in that there is an agent who reports a positive lie that is not the maximum value, that is a partial liar. For such an agent, the value to report represents the optimal trade off between the additional utility for reporting a higher value and the extra cost of reporting a more false value.

The behaviour of the fixed lie type in reporting a constant increment over the realised state from such a model is discussed in the work by Fischbacher and Föllmi-Heusi 2013, who argue that such a model is inconsistent with their results as it does not allow the significant over-reporting of the second highest value in the dice roll task as presented in their experimental results. For the model given here, however, this outcome represents an example of the one of extreme types of behaviour, whereas there are other behavioural types in the model that are consistent with their result.

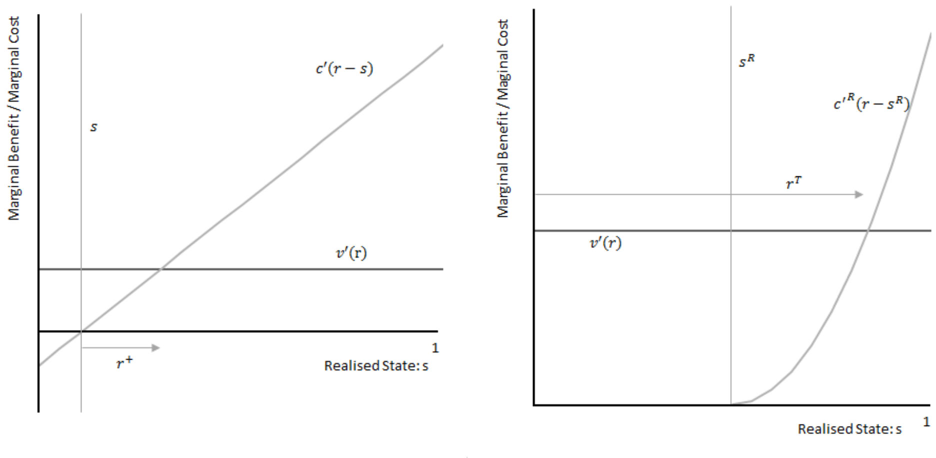


Figure 2.2: Illustration of first order conditions

The solution to the first order condition associated with the fixed lie type of agent is illustrated in the left hand plot of Figure 2.2. The optimal value for the agent to report can be considered as an increment, labelled r^+ in the figure, as this value gives the constant size of the lie the agent will choose over the set of states where the interior optimum continues to apply. Such an agent will never report a value less than the truth, giving rise to the previous observation that a model only containing a cost in terms of the size of the lie cannot produce negative lies.

The right hand plot of Figure 2.2 illustrates the term r^T previously referenced in the condition of the fixed value type. It represents the solution to the first order condition of the agent's problem when only the cost in terms of deviation from the reference point is present and holds as an interior optimum.

The six behavioural types presented so far reflect extremes of behaviour consistent with the model. A number of the types arise from the boundary conditions in the model, reflecting very small or very large moral considerations for lying. Others arise from considering only one of the cost terms in isolation, or from switching between the primary effect of the two cost terms. The following behavioural type seeks to characterise a more moderate form of agent who experiences both components of the lying cost.

Behavioural Type 7 *A moderate lying agent may exist that reports an increasing value with the state leading to positive lies in low states and negative lies for high states*

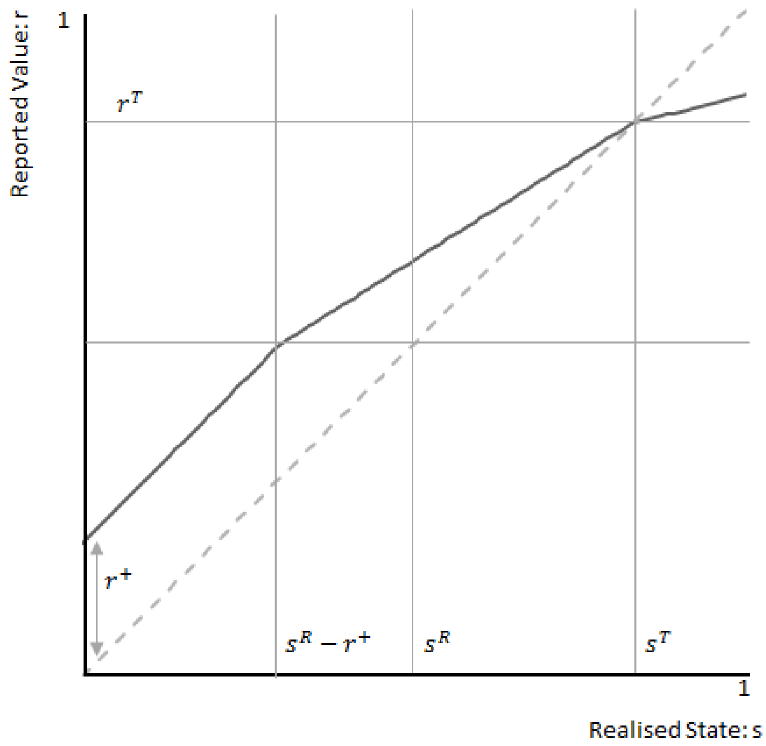


Figure 2.3: Illustration of value reported for a moderate lying type

A moderate lying agent experiences some cost in terms of the size of the lie, such that the type is willing to over-report their value, but only to some restricted degree. However, they also experience a cost in terms of the reference point, leading to higher costs for telling a lie of an equivalent size when the value reported lies above the reference point compared to below. Indeed, the additional cost for a particular lie above the reference point may be enough to cause the agent to switch to negative lies. A moderate lying agent is illustrated in Figure 2.3.

Figure 2.3 illustrates a number of features pertaining to a moderate lying agent. If the agent experiences sufficient cost in terms of the size of the lie that they will only over-report to a small extent in low states, then there will be some region where the agent exhibits behaviour consistent with a fixed increment. For any given state where the agent wishes to report a value below the reference point, the reference point will have no direct influence on the agent's choice, and the size of the increment of any lie will be constant. The highest state under which this consideration applies for can be determined from the first order condition, and is given by $s^R - r^+$, the state under which the agent will report a value equal to the reference point. If the agent experiences a sufficient cost in terms of deviation from the reference point then there may be some state where the agent will report the truth. The value of the state at which the agent will report the truth can also be determined from the appropriate first order condition, as will it occur when the agent reports the value r^T such that the state at which the agent reports the truth is s^T, where $s^T = r^T$. For such an agent, the extent of the lie reported between the states $s^R - r^+$ and s^T is falling. This change in behaviour reflects the influence of the cost of deviation from the reference point on the agent's behaviour. For realised states above the truth telling point the agent will no longer over-report the value and may switch to under-reporting.

A comparison of Figure 2.3 with the panels 2.1c and 2.1f of Figure 2.1 highlights the outcome of the combination of the two cost components on the behaviour of an agent. For realised states below the reference point, the cost in terms of the size of the lie may have the main effect driving behaviour where the agent reports a value with a fixed increment. For states above the reference point, the cost in terms of deviation from the reference point may be the main influence, producing behaviour akin to reporting some fixed value. The agent's behaviour for values in the central range, around the reference point, will reflect elements of both considerations. An agent who experiences both forms of the psychological cost contained in this model will therefore exhibit moderated behaviour in that they will typically over-report values for low realised states but reduce the degree of over-reporting for higher realised states and even under-report values for sufficiently high states. This is not to say this moderated behaviour applies to all agents, the model still allows for rational liars and honest types, only to agents who experience the appropriate relative levels of the two costs and the benefit from the lie.

A moderate lying agent is also consistent with the model of self-concept maintenance (Mazar et al. 2008). The agent is happy to over-report values to some extent, especially for low states. However the degree of this over-reporting will fall for such an agent as the state rises, eventually potentially turning into under-reporting of the value.

2.5 Change of Reference Point

The third proposition detailed that the value reported by an agent will be non-decreasing for an increase in the reference point. It can be noted that a number of the types described in the previous section will be unaffected by an increase in the reference point, in particular the fully dishonest, fully honest and fixed lie types of agent, who have no concern for the reference point or whose degree of concern is dominated by their cost in terms of the size of the lie. In all three cases proposition 3 holds and the value reported by the agent does not decrease with an increase in the reference point.

However other types may be affected by a change in the reference point and may increase the value they report for any given state upon an increase in the reference point. This outcome matches to the simple intuition that if the reference point is higher, such that the cost in terms of deviation from the reference point for a any given value for an agent to report affected by such a cost is lower, larger lies are more likely.

An example of a shift in the reference point is given Figure 2.4. The left hand panel of the figure illustrates how two different reference points, s^R and $s^{R'}$ where $s^{R'} > s^R$, can be extracted from the model of the previous section for a particular agent set over an extended theoretical range of states ($[-1, 1]$). The right hand panel shows the optimal response of the agent over the observable range $[0, 1]$ for the two reference points produced by overlaying the two sections taken from the left hand panel. The key feature illustrated in Figure 2.4 is that higher values are reported under the case of the higher reference point for the moderate lying agent under consideration. This agent's cost in terms of deviation from the reference point is such that the same degree of deviation is optimal, but this is applied relative to the reference point, such that the value of the state where the agent

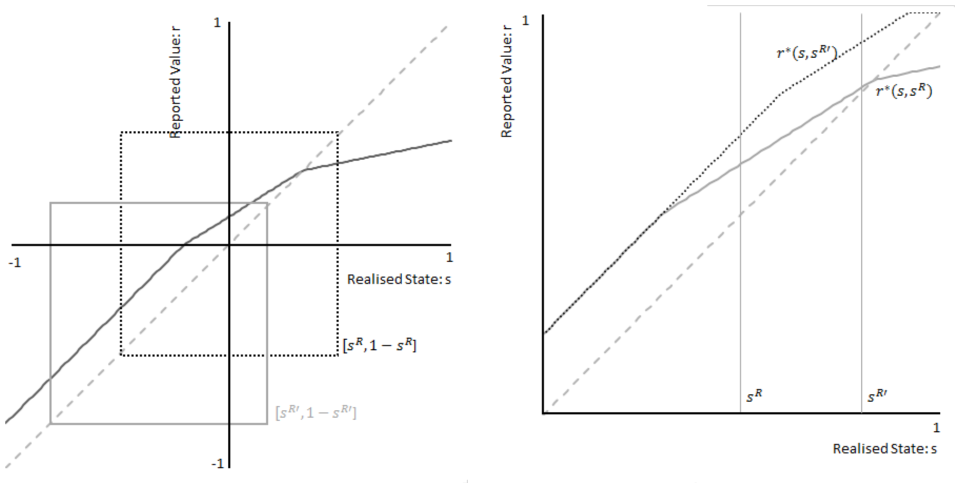


Figure 2.4: Illustration of value reported with different reference points

reports the truth (s^T) is greater under the case where the reference point is higher. In the case illustrated, the degree of change in the reference point is sufficient that under the lower reference point the agent will report negative lies in the highest states, whereas they report the maximum value for equivalent states under the higher reference point.

A further feature illustrated in Figure 2.4 is that there need not be an effect in behaviour upon a shift in the reference point. As noted at the beginning of this section a number of types consistent with the model will not be affected by a change in the reference point. However the illustration also shows that types that are affected by the change in the reference point may not be affected for all states. In particular where the cost of telling a lie in terms of the size of the lie is such that at the lower reference point the agent will report a value below the lower reference point, an increase in the reference point will have no effect.

2.6 Hypotheses

The approach to testing the model adopted in this chapter utilises a between subjects design in which each subject faces a single decision, with hypotheses based on the sample outcomes. An alternative test of the model would seek to examine the value an individual would report for all the possible values of the realised state under conditions with differing values of the reference point. Such a within subject design for an experiment, however, faces a problem of experimenter demand in eliciting valid responses from subjects being asked to respond multiple times, as described in Charness et al. (2012).

The propositions detailed in the previous sections allow for a number of hypotheses as a test of the model. The first two hypotheses are framed in relation to the value reported by subjects, whereas the latter two detail conjectures about outcome behaviours. The first hypothesis is a simple statement reflecting one of the key predictions of the model, that of partial lying.

Hypothesis 1 *A proportion of the sample will report false values that are not the maximum value*

The first hypothesis states that some individuals will report a partial lie, that is a false value that is not the maximum value. A response that is a partial lie is consistent with a number of the types discussed in the previous section, such as the fixed lie type (for appropriate values of the realised state).

The second hypothesis details a prediction about the value that will be reported under two separate conditions with different reference points, such that $s^{R'} > s^{R}$. The hypothesis proposes a pattern of change in the reported value upon a shift to the higher reference point. Proposition 3 gives that the reported value is non-decreasing upon an increase in the reference point, such that a simplistic hypothesis would conjecture that there would be no decrease in the value reported upon an increase in the reference point. A hypothesis of this form would, however, offer no evidence in support of the role of the reference point in the model. To create a hypothesis that is able to provide positive evidence requires an analysis relating to how any of the types of agents consistent with the model would be affected by an increase in the reference point. To support this analysis, the hypothesis splits the decision space into four quadrants based on the application of lower of the two reference points to both the dimension of the realised state and of the reported value. The resulting quadrants can be labelled by the appropriate corresponding compass points. The SW quadrant applies for values of the realised state and the reported value are below the lower reference point s^R ($s < s^R$, $r < s^R$), such that the NW quadrant applies when $s < s^R$ and $\geq s^R$ and the NE quadrant where both the value of the realised state and the reported value are on or above s^R ($s \geq s^R, r \geq s^R$). The theory predicts that no values should be reported in the SE quadrant.

Hypothesis 2 *There should be no change in the value reported by individuals reporting in the SW quadrant under the lower reference point upon an increase in the reference point. The value reported by some individuals in the NE and NW quadrants will be increased*

The utility function of the model states that the cost in terms of lying has no impact when an agent reports a value below the reference point, such that the reference point has no influence on values reported in the SW quadrant. As this condition will still apply under an increase in the reference point, an agent's choice will not be changed, leading to the first part of the hypothesis. To detect positive evidence in favour of the model presented here with a reference point therefore requires consideration of the NW and NE quadrants, where the value reported is on or above the lower reference point (under the condition of the lower reference point). While there will be some agents in these regions unaffected by a shift in the reference point (such as characterised by the Honest or Dishonest types), if there are agents of the types affected by the reference point, then the increase in the reference point will lead to an increase in the value reported for a given realised state in the NW and NE quadrants.

The remaining hypotheses consider the propensity to report the maximum value or the truth under the conditions of the different reference points.

Hypothesis 3 *The proportion of the sample reporting the maximum value will increase in the treatment with a higher reference point for realised states in excess of* s^R

The third hypothesis reflects that there is an upper bound to the reporting decision, such that if the previous hypothesis that agents will report increased values under a condition with a higher reference point holds, then a greater proportion of agents will report the maximum value. Gneezy et al. (2013) report a decline in the propensity to report the maximum value with an increase in the realised state in a sender-receiver game. The authors of that study state that the pattern observed is inconsistent with a simple model of a convex cost of lying. As previously noted, Fischbacher and Föllmi-Heusi (2013) also state that their results from a dice rolling task are inconsistent with a simple model of a convex cost of lying. The inclusion in the model of an additional cost in terms of deviation from the reference point allows for a decline in the propensity to report the maximum value and furthermore the hypothesis that the proportion of subjects reporting the maximum value will be greater under the condition of the higher reference point.

Hypothesis 4 *There will be a greater propensity for truth telling in the treatment with the lower reference point for states in excess of* s^{R}

The fourth hypothesis conjectures that if there is less lying expected in the condition with the lower reference point, as given in the previous two hypotheses, then there will be more truth telling as a result.

2.7 Experimental Design and Procedures

The experimental design is based on a numeracy test of sufficient difficulty that the majority of subjects were not able to correctly answer all questions in the allocated time. Cheating was possible in the experiment as subjects were required to pay themselves for their performance before turning in their assessment booklets along with the unclaimed proportion of their potential payment as they left the laboratory.

Invitations to laboratory sessions were sent to potential subjects from a pool of previously registered students at the university of Exeter using the ORSEE system (Greiner 2015). Nine sessions were held over three separate days between the 24 January and 31 January 2017. Thirty two subjects were invited to each of the nine sessions, with an average show up rate of 89.6%.

Before each session commenced, the thirty two desks in the laboratory were pre-configured with the experimental materials. The set of items placed on each desk was identical and consisted of a large clear plastic bag, a small white envelope containing twenty 20 pence coins, three pound coins and two white A4 envelopes, one marked with a letter A, the other marked with a letter B. An image of the configuration of one of the desks is given in Figure B.1 in the Appendix. Each of the desks was clearly labelled with a desk number, ranging from 1 to 32.

Upon entry to the laboratory, each subject was given a participant payment receipt form, requested to draw a token revealing a number in the range 1-32 and invited to sit at the desk labelled with the number drawn. Subjects were asked to fill in their name on a blank payment form and sign it before the experimental tasks began.

Subjects were then requested to take the booklet out of envelope A and the instructions from the first page were read aloud. Full instructions for the experiment are given in the Appendix. The instructions began that there was a £3 show up fee for the experiment, and there may be an opportunity for subjects to make more money through their decisions in the experiment. The instructions detailed that subjects should not use their name or any other identifier on the lab materials other than the payment form that they had already signed to ensure the anonymity of responses. Subjects were informed that the experimenters would not examine the payment receipts and that these were to be processed only by the university's finance team.

The instructions further stated that the experiment would be in two parts, the first of which was a series of questions in a booklet attached to the instructions sheet. Subjects were given ten minutes to complete the booklet, which contained a section on personal characteristics, a question relating to the subject's risk aversion preferences, an inter-temporal investment decision and a cognitive reflection test.

At the end of the ten minutes, subjects were asked to put the booklet for the first part of the experiment to one side and open the second large A4 envelope marked with the letter B. Again the instructions from the first page of the booklet were read out loud. The instructions for the second part of the experiment detailed a numeracy test. Subjects were told that they were required to complete a series of calculations of the form 3+5-2=. The instructions stated that subjects would be given thirty seconds to complete as many out of a total of twenty calculations as possible and that each solution would earn 20 pence. The instructions described that at the end of the thirty seconds subjects were to check how many calculations they had answered and pay themselves accordingly from the envelope of 20 pence pieces on the desk. Subjects were then required to place the two booklets for the experimental tasks and the envelope containing any remaining coins into the clear plastic bag and seal it. Subjects were also asked to complete the payment form with the total amount for the experimental session, but to keep that separate.

The instructions stated that upon completion of these tasks, subjects would be free to leave, but were requested to place the clear plastic bag containing experimental materials in the left hand box of two boxes, and the payment receipt face down in a separate box by the door of the laboratory as they left.

Once the instructions had been read out, subjects were given thirty seconds to complete the numeracy task, whereupon the instructions for processing the experimental materials were repeated, the subjects were thanked for their time and allowed to leave the laboratory.

Three initial sessions were run as a CONTROL. From these sessions, two values were found for the most common response by the subjects (measured in the manner described in the next paragraph), with 18% of subjects reporting a value of ten and a matching 18% reporting a value of twenty. These modal responses from the CONTROL were selected for use in the treatment sessions. Three further sessions were then performed for each of two additional treatments (leading to a total of nine sessions) using one of the two modes from the initial control treatment to set a reference point. In the first treatment, LOW, an additional line was added to the instructions that stated:

> In previous sessions using this test one of the two most commonly reported values was 10 questions.

In a second treatment, HIGH, an alternative line was added to the instructions that stated:

> In previous sessions using this test one of the two most commonly reported values was 20 questions.

Upon completion of the experiment, the plastic envelopes returned by each of the subjects were opened and the contents inspected. The number of coins returned was counted and recorded and an implied figure of twenty minus the number of coins returned was calculated as the score reported by a subject. The numeracy test was marked and the number of questions attempted by a subject as well as the number of solutions that were correct were recorded. Answers from the booklet as to a subjects' gender and their other choices made up the complete base data set used.

The average earnings for the experiment were £5.79, for a task that lasted 30 minutes. The treatments are summarized in Table 2.1

Treatment	Subjects	% Male	Description
CONTROL	88	36.6	No reference to previous sessions
LOW	82	47.6	Mode of CONTROL revealed as 10
HIGH	88	50.0	Mode of CONTROL revealed as 20

Table 2.1: Treatment overview

2.8 Results

2.8.1 Variables

The following discussion uses a number of terms to describe values recorded from the experiment. The first is ReportedCalculations, which is the number of calculations a subject paid themself for, as implied from the number of coins they returned at the end of the session. A subject returning no coins therefore has a value for ReportedCalculations of 20. The second term is CorrectCalculations, which is the number of calculations on the answer sheet a subject answered correctly, as checked by the experimenter after the sessions. A third measure, AttemptedCalculations, is the number of calculations a subject wrote an answer for on the experimental booklet, whether the answer was correct or not.

Responses by subjects whereby the value of ReportedCalculations is greater than the number of AttemptedCalculations, that is when subjects paid themselves for calculations they did not attempt, are classified as reporting a Lie. Responses by subjects where the value of ReportedCalculations is equal to or less than the number of AttemptedCalculations but the value of ReportedCalculations is greater than the number of CorrectCalculations are classified as a Misreport. This latter category therefore covers responses that may arise due to deliberate cheating by subjects or may arise due to genuine error in performing one or more calculations. All responses where the value of ReportedCalculations is greater than the number of CorrectCalculations are classified as a False report, encompassing both the Lie and Misreport classifications.

A response classified as a Lie is an explicit act of cheating, a subject behaving in this manner took money for questions they had attempted. The forms of behaviour underlying reports classified as Misreports are, however, more complex, as the number and nature of the incorrect answers found varies. For instance, in some cases it is enough that a subject answered a single question incorrectly, most probably in error, for their response to be classified as a Misreport. However, a response whereby a subject wrote the same value for all the answers, which would include a large number of incorrect solutions, would also be classified as a Misreport. Cases with a moderate proportion of incorrect answers are also classified as a Misreport, though the intended behaviour of the subject may be unclear. The observation of such behaviour has a parallel in studies of tax evasion, where it is not always clear if non-compliance is a deliberate act or a result of error (Andreoni et al. 1988).

In order to test the second hypothesis, the reporting space is compartmentalised into four quadrants by a series of dummy variables relating to the value of the reference point set in the LOW treatment. The first of these dummy variables, CCAR, takes a value of 1 when the number of CorrectCalculations is equal to or greater than 10, otherwise the value is 0. The second, RCAR, takes the value 1 when ReportedCalculations is equal to or greater than 10, otherwise the value is 0. These variables allow for specification of four regions within the reporting space. The NE quadrant of the reporting space, as detailed in the hypotheses, therefore refers to the region where CCAR=1 and RCAR=1.

The binary variable Male takes a value of 1 if a subject reported being male and 0 otherwise. The term Age refers to the age reported by the subjects. The variable Allocation is the subject's choice for investment in the risky component in the risk aversion measure of the first part of the experiment, Option details their choice in the an inter-temporal investment decision and CRTCorrect is a binary variable that takes the value 1 if the subjects response on the cognitive reflection test was the correct value. The terms Extroversion, Agreeableness, Emotional Stability, Conscientiousness and Openness refer to answers given on a scale from 1 to 7 about the subjects degree of agreement with two statements assessing their personality for each of these traits (Woods and Hampson 2005). The pairs of statements used for this assessment are given in the Appendix. The variables Economics, AccountingFinance and STEM take a value of 1 if a subject stated that their course of study was economics, accounting and finance or part of science, technology, engineering and maths respectively.

2.8.2 Methods

Non-parametric tests are two-sided unless stated as one-sided (1S) in relation to a directional hypothesis. The primary response variable is ReportedCalculations, which is censored on the right hand side by the maximum score on the numeracy test of 20. The value for ReportedCalculations is also technically censored on the left hand side at zero, but as no subject reported fewer than 5 calculations, left censoring has no effect. Censored regressions were performed using a limiting value of 20 for the dependent variable reflecting the maximum possible score on the numeracy test (Tobin 1958). The declaration by each subject (indexed by i) is treated as an independent response. The censored regression model for individual i can be stated as a latent variable in relation to a vector of independent variables ($\mathrm{x_i}$) as:

$$
\begin{gathered}
\mathrm{ReportedCalculations}_i^* = \mathbf{x_i'}\boldsymbol{\beta} + \mathrm{u}_i \\
\mathrm{ReportedCalculations}_i = \mathrm{ReportedCalculations}_i^* \quad \text{if} \quad \mathrm{ReportedCalculations}_i^* < 20 \\
\mathrm{ReportedCalculations}_i = 20 \quad \text{if} \quad \mathrm{ReportedCalculations}_i^* \geq 20 \\
\mathrm{u}_i \sim \mathrm{N}(o, \sigma^2)
\end{gathered} \tag{2.14}
$$

where the primary vector of explanatory variables consists of the main treatment terms, the variables to separate the reporting space and the set of appropriate interactions, as given by the following:

$$
\begin{gathered}
\mathbf{x_i'}\boldsymbol{\beta} = \beta_0 + \beta_1\mathrm{RCAR}_i + \beta_2\mathrm{CCAR}_i + \beta_3\mathrm{LOW}_i + \beta_4\mathrm{LOW}_i * \mathrm{RCAR}_i \\
+ \beta_5\mathrm{LOW}_i * \mathrm{CCAR}_i + \beta_6\mathrm{HIGH}_i + \beta_7\mathrm{HIGH}_i * \mathrm{RRAR}_i + \beta_8\mathrm{HIGH}_i * \mathrm{CCAR}_i \\
+ \beta_9\mathrm{CorrectCalculations}_i + \beta_{10}\mathrm{RCAR}_i * \mathrm{CorrectCalculations}_i \\
+ \beta_{11}\mathrm{CCAR}_i * \mathrm{CorrectCalculations}_i + \beta_{12}\mathrm{LOW}_i * \mathrm{CorrectCalculations}_i \\
+ \beta_{13}\mathrm{LOW}_i * \mathrm{RCAR}_i * \mathrm{CorrectCalculations}_i + \beta_{14}\mathrm{LOW}_i * \mathrm{CCAR}_i * \mathrm{CorrectCalculations}_i \\
+ \beta_{15}\mathrm{HIGH}_i * \mathrm{CorrectCalculations}_i + \beta_{16}\mathrm{HIGH}_i * \mathrm{RRAR}_i * \mathrm{CorrectCalculations}_i \\
+ \beta_{17}\mathrm{HIGH}_i * \mathrm{CCAR}_i * \mathrm{CorrectCalculations}_i
\end{gathered} \tag{2.15}
$$

The coefficients of the regression relating to Equation 2.15 are the partial effects with relation to the conditional mean of the latent variable. Regressions based on Equation 2.15 are referred to as Model 1. Model specification tests using each of the variables corresponding to responses in the questionnaire from the first part of the experiment revealed a significant correlation between the subject's choice and their gender. A second regression model (Model 2) adds an additional regression term to Equation 2.15 for the variable Male. A third regression model (Model 3) adds a term to the regression for each of the variables detailed in the previous section.

A test for heteroskedasticity of the regression results revealed a correlation between the generalised residuals of the censored regression and the value of CorrectCalculations (Chesher and Irish 1987; Pagan and Vella 1989). An additional model (Model 4) uses an alternative form for the variance given by:

$$V\{u_i\} = \sigma^2\{exp(\alpha_0 + \alpha_1 * CorrectCalculations_i)\}^2 \tag{2.16}$$

Model 5 utilises the heteroskedastic form of the variance given in Equation 2.16 and adds to variable Male to the regression Equation 2.15.

A multinomial logistic model can be used to examine a classification of subjects' reports (Nerlove and Press 1973). Let y_i denote the classification of the choice made by the subject i with some associated vector of observable characteristics $\mathbf{x_i}$. The probability that a subject chooses a particular classification, j, from the a set of classifications $0, 1 \ldots J$ is given by $P(y_i = j|\mathbf{x_i})$. The multinomial logistic model for the subjects choice can be expressed as the following:

$$P(y_i = j|\boldsymbol{x}_\mathbf{i}) = \frac{exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_j})}{\sum_{k=0}^{J} exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_k})} \tag{2.17}$$

To remove the indeterminacy in the model, a normalisation was conducted for one of the subjects' choices (Correct) such that the values are zero ($\beta_0 = 0$). The expression for the model can be rewritten as:

$$P(y_i = j|\boldsymbol{x}_\mathbf{i}) = \frac{exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_j})}{1 + \sum_{k=1}^{J} exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_k})} \tag{2.18}$$

The base expression for the independent variables in the regression is formed from the interactions of the value of CorrectCalculations and the treatment:

$$x_i'\boldsymbol{\beta}_j = \beta_{j0} + \beta_{j1}\mathrm{CorrectCalculations}_i + \beta_{j2}\mathrm{LOW}_i + \beta_{j3}\mathrm{HIGH}_i + \\ \beta_{j4}\mathrm{CorrectCalculations}_i * \mathrm{LOW}_i + \beta_{j5}\mathrm{CorrectCalculations}_i * \mathrm{HIGH}_i \quad (2.19)$$

The model given in Equation 2.19 is referred to as Model L1. An extended model for the regression further incorporates other variables relating to the personal characteristics of the subjects (Model L2). Greene (2010) proposes a mechanism for addressing issues relating to the usage of hypotheses tests of the coefficients of interaction effects in non-linear models raised by Ai and Norton (2003). The approach to analysis suggested by Greene (2010) is to perform tests relating to specification of the model prior to analysis, for which the author argues that graphical presentations create an informative adjunct to any statistical analysis. I adopt this recommended approach in the analysis of the results. The results presented make extensive use of the post-estimation calculated propensities at the integer values of CorrectCalculations. An overview of the approach used for the analysis is given in Williams (2012). Results for the regressions performed are presented in the Appendix.

2.8.3 Overview

Figure 2.5 displays the mean number of CorrectCalculations recorded in each of the three treatments. There is no difference in the number of CorrectCalculations between the HIGH and LOW treatment (Mann-Whitney test (henceforth MW) $z = 0.003$, $p = 0.9975$) though there is somewhat higher average number of CorrectCalculations observed in the HIGH (MW: $z = 1.705$, $p = 0.0882$) and LOW (MW: $z = 1.65$, $p = 0.0987$) treatments compared to the CONTROL, which suggests that there may have been subjects with a greater degree of numeric ability in these sessions. The number of subjects correctly answering twenty questions, however, did not vary between the treatments (Fishers Exact Test (henceforth FET): $p = 1.000$) with 5.68%, 4.88% and 4.55% of subjects answering all the

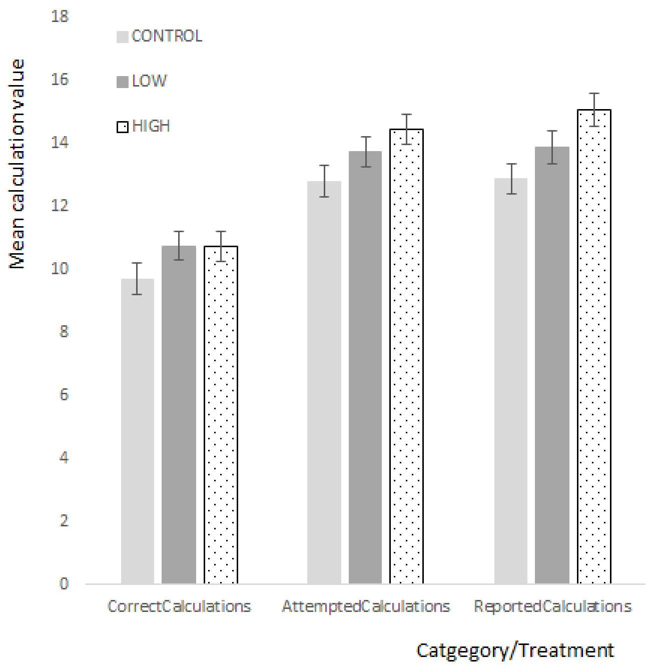


Figure 2.5: Mean values of CorrectCalculations, AttemptedCalculations and ReportedCalculations

questions correctly in the CONTROL, LOW and HIGH treatments respectively. Crucially, the number of CorrectCalcualtions did not fall and the observation of slightly higher correct scores in the LOW and HIGH treatments compared to CONTROL indicating that the presence of an additional sentence in the instructions relating to the scores of previous subjects did not negatively affect subject's performance on the test.

Figure 2.5 shows no difference in the number of AttemptedCalculations between the CONTROL and LOW treatments (MW: $z = 1.53$, $p = 0.1228$) or between the LOW and HIGH treatments (MW: $z = 1.072$, $p = 0.2836$), however there is a difference observed between the CONTROL and HIGH treatments (MW: $z = 2.501$, $p = 0.0124$). The figure also shows a higher number of ReportedCalculations in the HIGH treatment compared to the LOW treatment (MW: $z = 1.821$, $p = 0.0686$) and the CONTROL treatment (MW: $z = 3.131$, $p = 0.0017$), though no difference between the LOW and CONTROL treatments (MW: $z = 1.598$, $p = 0.1099$).

Figure 2.6 shows bubble plots of the number of ReportedCalculations against the number of AttemptedCalculations (top row) and the number of CorrectCalcu-

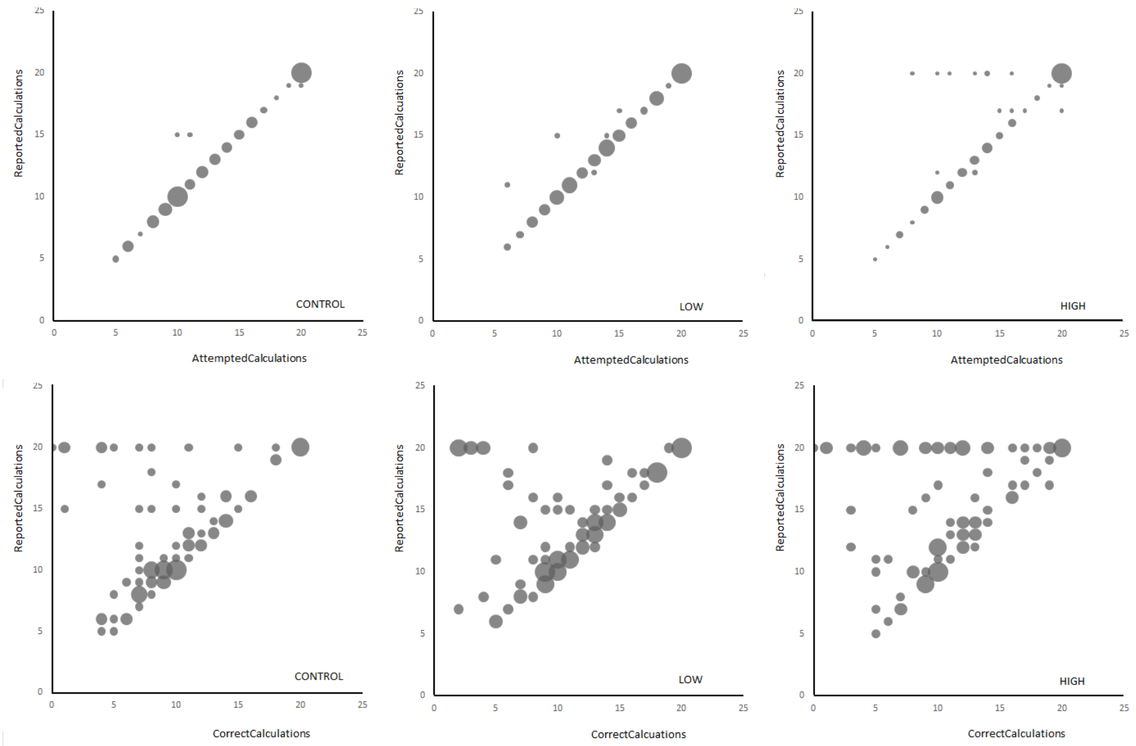


Figure 2.6: Bubble chart of ReportedCalculations by AttemptedCalculations and by CorrectCalculations

lations (bottom row) for the three treatments. Lies can be identified by off-diagonal bubbles in the plots in the top row, Misreports can be observed in the off-diagonal bubbles in the plots in the bottom row.

The experimental procedures gave subjects an additional mechanism through which they could cheat, in that at the end of allowed time for performing the calculations when they were asked to count up how many questions they had completed and pay themselves, subjects could have used some of this time to answer additional questions. Cheating through such a manner would not be identified within this experiment. To this end, the experimenter remained in the room throughout the experiment and no cases of subjects continuing to undertake calculations were observed. The plots in Figure 2.6 also give evidence for little or no cheating through this mechanism, as such behaviour would have lead to a higher proportion of responses in the top right hand corner of the plots.

2.8.4 Reported Value

Result 1 *A proportion of the sample reported False values that were not the maximum value*

The first hypothesis from the model predicts that some subjects will report a False value that is not the maximum value. A comparison of the upper plot against the lower plot for each column of Figure 2.6 reveals a markedly higher proportion of subjects whose responses are classified as a Misreport compared to those that are a Lie for all three treatments. 64.7% of all subjects' responses were classified as False, where 6.2% of responses were classified as a Lie and a further 58.5% were classified as a Misreport. False values reported by subjects that are not reports of the maximum value, can be identified as bubbles in the lower panels of Figure 2.6 that lie above the diagonal but blow the line where ReportedCalculations is equal to 20. 46.5% of all subjects were observed to report values lying in this region.

Result 2 *A higher level of False values were reported by subjects in the HIGH treatment compared to the LOW treatment in the NW and NE quadrants of the reporting space*

The second hypothesis predicts a change in behaviour in reporting False values between the LOW and HIGH treatments in relation to the shift in the reference point. There is no significant difference in the proportion of subjects reporting a False value between the treatments (FET: p = 0.847), however Figure 2.6 clearly shows differences in the subjects' responses between the HIGH and LOW treatments. The results of the censored regression detailed in Equation 2.15 are given in Table B.1 in the Appendix and illustrated in Figure 2.7, which shows the estimated regression plots for the LOW and HIGH treatments in the three key quadrants of the reporting space. Results for the statistical tests of the joint hypotheses that both the intercept and slope are equal between the LOW and HIGH treatments in each of the quadrants are illustrated in Figure 2.7 and detailed in Table 2.2. As noted in the methods section, there was a considerable

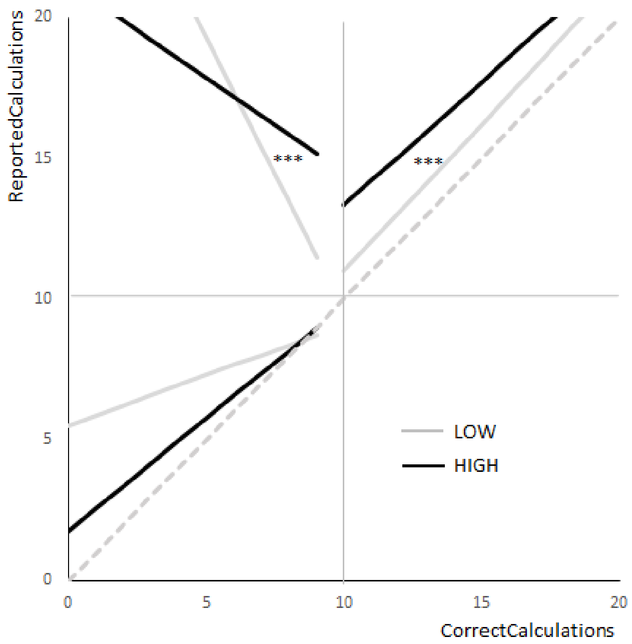


Figure 2.7: Illustration of results for censored regression of CorrectCalculations on ReportedCalculations

2

degree of heteroskedasticity found in the regression results. As the presence of heteroskedasticity will bias the estimators calculated by a censored regression (Pagan and Vella 1989), a further regression with an additional term to represent the heteroskedasticity was also performed. The results, given in Table B.1 in the Appendix are not significantly altered.

Quadrant	Region	Model 1	Model 3
SW	CorrectCalculations <10, ReportedCalculations <10	p=0.7342	p=0.6868
NW	CorrectCalculations <10, ReportedCalculations ≥ 10	p=0.0097 ***	p=0.0227 **
NE	CorrectCalculations ≥ 10, ReportedCalculations ≥ 10	p=0.0090 ***	p=0.0423 **

Table 2.2: Results for joint test of difference of intercept and slope

Figure 2.7 illustrates two key points in relation to the second of the hypotheses. The first is that there is no significant difference in behaviour between the LOW and HIGH treatments in the SW quadrant, that is for subjects with a value of less than 10 for both CorrectCalculations and ReportedCalculations. The bubble plots of Figure 2.6 illustrate a pattern of a mixture of correct reporting and small levels of misreporting in the SW quadrant of the reporting space consistent with the regression results of Figure 2.7. The second key point is that there is a significant difference in the value reported between the treatments in both the NW and NE quadrants.

In the NW quadrant, the high intercepts and negative slopes observed in both treatments arise from a proportion of subjects who report the maximum value for very low values of CorrectCalculations combining with those who report slightly above the number of CorrectCalculations for a value of CorrectCalculations just under 10. The less negative slope found for the HIGH treatment relative to the LOW treatment in the NW quadrant indicates that there is more Misreporting as the value of CorrectCalculations rises to 10 in the HIGH treatment compared to the LOW treatment. For the NE quadrant, the larger intercept and lower slope calculated for the HIGH treatment compared to the LOW treatment are indicative of larger false values being misreported in the HIGH treatment.

Table 2.3 shows that the proportions of the sample observed in each of the quadrants are very similar between the LOW and HIGH treatments, indicating that the patterns observed in the regression are due to differences in behaviour within the quadrants rather than arising due to subjects moving responses between the quadrants. The values for the CONTROL treatment are shifted slightly to the SW and NW quadrants, reflecting the slightly lower number of CorrectCalculations under this treatment, as demonstrated in Figure 2.5.

Quadrant	CONTROL	LOW	HIGH
SW	23.9%	14.6%	11.4%
NW	29.5%	26.8%	26.1%
NE	46.6%	58.5%	62.5%

Table 2.3: Proportion of sample observed in reporting quadrants

A generalised residuals test for misspecification (Chesher and Irish 1987; Pagan and Vella 1989) indicates that the subject's gender does have a significant role in the regression. Table B.1 shows that differences reported are reasonably robust to the addition of gender and the other information recorded about the subjects. A further test for heteroskedasticity also revealed an issue with residuals found to be correlated to the value of CorrectCalculations. Table B.1 further shows that the results are robust to an extended censored regression using the expression given in 2.16 to model the heteroskedasticity.

2.8.5 Classifications

The final two hypotheses propose expected behaviour in relation to the reporting of particular values, namely the maximum value and the correct value. Subjects' responses were classified into one of a number of types relating to if they reported the maximum value, the correct value or some value in between. Details of the classifications used are given in Table 2.4, which shows that the categories of Lie and Misreport described in the previous section could be further subdivided according to subjects who reported the maximum value falsely (that is ReportedCalculations=20 but CorrectCalculations$<$20), whereby the response is classified with the prefix Full and those who reported a false value that was not the maximum value, which are classified with the prefix Partial. Responses by subjects who reported the number of questions they had solved correctly on the numeracy test are classified as Correct and the small number of subjects (1.2%) who reported a value less than the number they solved correctly are classified as an UnderReport.

Label	Description
FullLie	ReportedCalculations=20 $>$ AttemptedCalculations
PartialLie	20 $>$ ReportedCalculations $>$ AttemptedCalculations
FullMisreport	ReportedCalculations=20 = AttemptedCalculations $>$ CorrectCalculations
PartialMisreport	20 $>$ ReportedCalculations $\leq$ AttemptedCalculations; ReportedCalculations $>$ CorrectCalculations
Correct	ReportedCalculations = AtteptedCalculations = CorrectCalculations
UnderReport	ReportedCalculations $<$ CorrectCalculations

Table 2.4: Response Classifications

In terms of False reporting, the classifications of FullLie or a FullMisreport can be combined to form a further classification of FullFalse. Similarly, the classifications of PartialLie and PartialMisreport can be combined to form a classification of PartialFalse reports.

Figure 2.8 shows the proportion of subjects' responses by classification for each of the treatments. The following subsections detail a number of the differences in the proportions reporting by classification between the treatments visible in the Figure. As the proportion of subjects found to be under-reporting is small and the extent to which those under reporting did so is negligible, responses originally classified as an UnderReport are re-classified as Correct in the following discussion.

Result 3 *A higher proportion of subjects where observed to report a FullLie in the HIGH treatment compared to the LOW treatment*

6.2% of all subjects were classified as reporting a Lie, with a significantly higher level of Lies found in the HIGH (11.4%) treatment compared to the CONTROL (2.3%) treatment (FET: p = 0.032), though not compared to the LOW (4.9%) treatment (FET: p = 0.165). A significantly higher proportion of subjects (8.0%) were observed to have reported a FullLie in the HIGH treatment compared to there being no subjects seen to behave in this manner in the LOW treatment (FET:1S p = 0.009) and the CONTROL treatment (FET: p = 0.014).

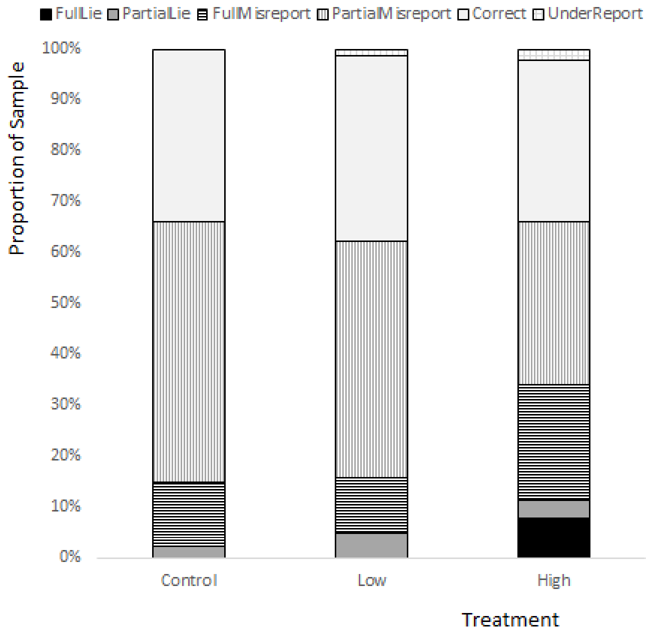


Figure 2.8: Proportion of subjects reporting in each of the classifications

Result 4 *A higher proportion of subjects where observed to report a FullMisreport in the HIGH treatment compared to the LOW treatment*

There are no differences in the proportion of subjects' responses categorised as a Misreport between the treatments (FET: p = 0.443). There is, however, a significant difference in the proportion of subjects' responses classified as a FullMisreport, with 22.7% of subjects in the HIGH treatment reporting in this manner compared to 11.0% in the LOW treatment (FET:1S p = 0.033). The observation of higher proportions of subjects' responses that are a FullLie or a FullMisreport in the HIGH treatment compared to the LOW treatment leads to the conclusion that there will be a higher proportion of FullFalse values in the HIGH treatment, detailed by the following result.

Result 5 *A higher proportion of subjects reported a FullFalse value in the HIGH treatment compared to the LOW treatment. The calculated propensity to report a FullFalse value is greater in the HIGH treatment for values of CorrectCalculations above 7*

30.7% of subjects reported a FullFalse value in the HIGH treatment compared to 11.0% in the LOW treatment (FET:1S p = 0.001). The results of the multino-

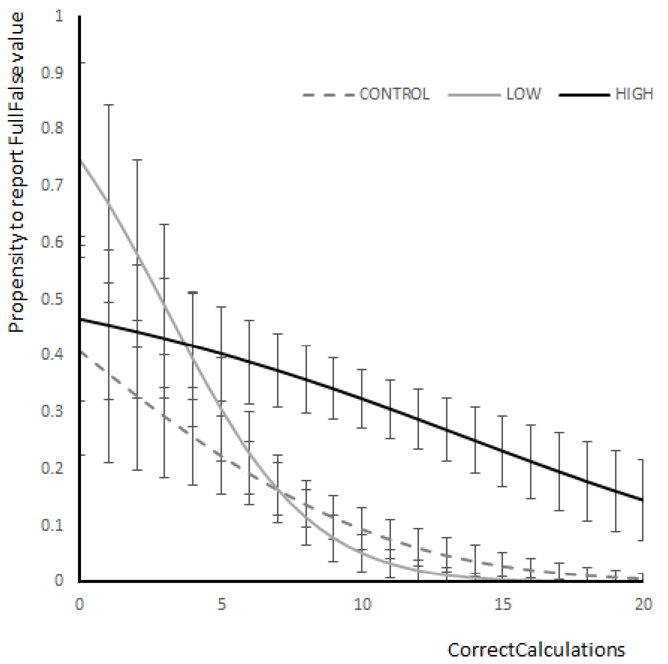


Figure 2.9: Propensity of subjects to report a value classified as FullFalse

mial logistic regression detailed in Equation 2.19 for the classifications FullFalse, PartialFalse and Correct are given in Table B.3 in the Appendix. The calculated propensity for subjects to report a FullFalse value by the level of the CorrectCalculations for each of the treatments is illustrated in Figure 2.9. Statistical tests for the contrast in propensity are given in Table B.4 in the Appendix. The results show that there is a higher propensity to report a FullFalse value in the HIGH treatment compared to the LOW treatment for a number of CorrectCalculations in excess of 7.

The proportion of subjects reporting a Correct value is similar in the CONTROL treatment (34.1%) to the LOW treatment (36.6%) and the HIGH treatment (31.8%), and the difference between the LOW and HIGH treatments is not significant (FET:1S p = 0.311). This finding is evidence against the fourth hypothesis. Figure 2.10, however, illustrates there is a higher propensity to report a Correct value in the LOW treatment compared to the HIGH treatment for values of CorrecCalculations in excess of 16. Table B.5 in the Appendix shows results for statistical tests of the contrast in propensity. This result suggests that subjects' behaviour may be consistent with the fourth hypothesis, but only for high values

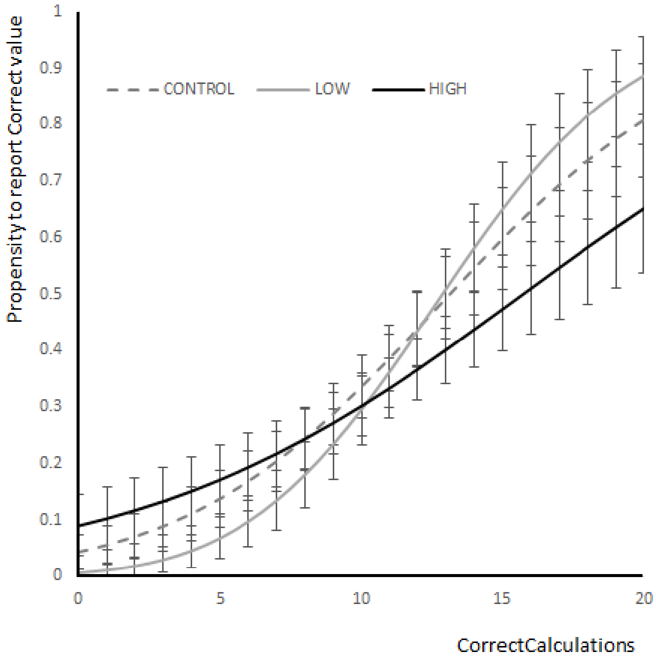


Figure 2.10: Propensity of subjects to report a value classified as Correct

of CorrectCalculations that mean the overall sample test does not report them.

A similar proportion of responses made by subjects were classified as a PartialFalse value in the CONTROL (53.4%) and LOW (51.2%) treatments. A lower proportion of the responses made by subjects in the HIGH treatment (35.2%) were classified as a PartialFalse value (FET: p = 0.044 compared to LOW treatment). Figure 2.11 illustrates the calculated propensity to report a PartialFalse value by the number of CorrectCalculations in the three treatments. There is an increased propensity to report a PartialFalse value in the LOW treatment compared to the HIGH treatment for scores of CorrectCalculations between 5 and 13, as demonstrated by statistical tests of the contrast in propensity shown in Table B.6 in the Appendix.

2.8.6 Behaviour

A key feature of the results is the difference between the proportion of subjects whose responses were classified as a Lie (6.2%) the those classified as a False value (64.7%), the difference being the proportion of responses classified as a Misreport. The bottom row of panels of the bubble plots presented in Figure

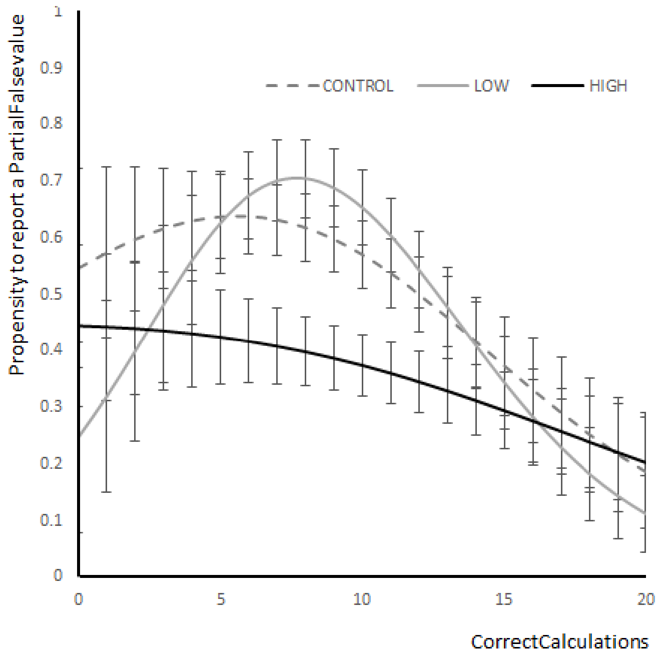
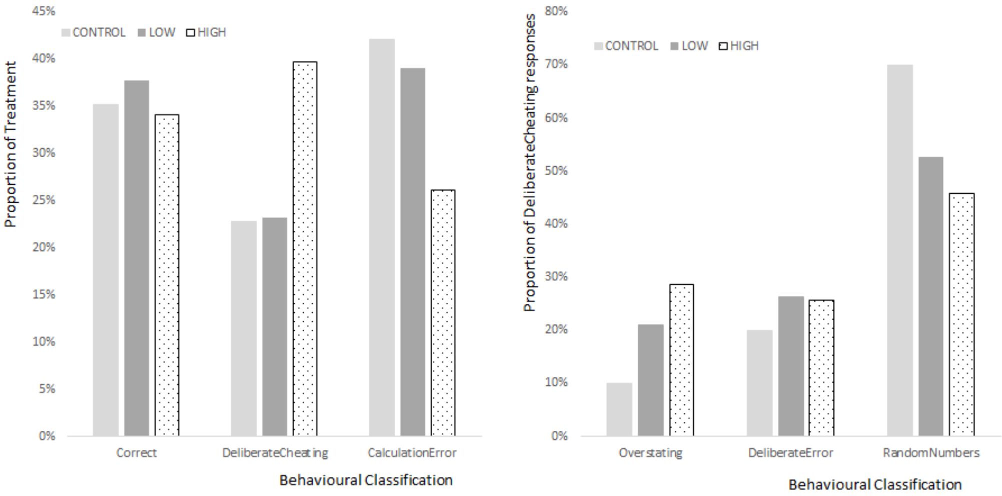


Figure 2.11: Propensity of subjects to report a value classified as PartialFalse

2.6 suggest a number of different types of behaviour by subjects who were classified as misreporting, in particular some subjects reported highly inflated values, others reported intermediate values while others only misreported by a small amount. The following section seeks to further examine the nature of misreporting by the subjects.

The answer sheets submitted by subjects were coded independently by two experimenters according to the behaviour used to complete the experiment. Three major categories were identified, the first of which, DeliberateCheating, was comprised of a number of minor categories. Subjects were judged to have deliberately cheated in one of three ways. The first mechanism identified for cheating was simple overstating of the number of calculations attempted (Overstating), correlating to the reporting a Lie in the previous sections. The second mechanism was the use of random numbers not related to the actual calculations being undertaken (RandomNumbers). The third mechanism was to make a deliberate error, typically in the form of a single response to calculations, or a incremental sequence used to complete the solutions sheet (DeliberateError). The second major behavioural category identified was Correct, whereby subjects reported the number



(a) Proportion of sample categorised by major behaviour

(b) Proporion of subjects regarded DeliberateCheating categorised by minor behaviour

Figure 2.12: Behavioural characterisation of answer booklets

of calculations they had answered correctly. The third major category identified was CalculationError, whereby subjects answered a small number of calculations incorrectly, typically with values close to the actual correct answer. In a small number of cases the categorisations by the experimenters differed, in particular between DelibderateCheating through RandomNumbers and CalculationError. In such cases, the results presented err on the side of caution and use the category of CalculationError for such responses.

Result 6 *The proportion of subjects' answer sheets categorised as DeliberateCheating is greater in the HIGH treatment than the LOW. The proportions of subjects reporting by each of the three identified mechanisms for cheating are the same between the HIGH and LOW treatments*

Figure 2.12a shows there are no differences in the proportions of responses assigned to the Correct category, in keeping with the results presented in the previous section. It is also important to check the HIGH treatment does not have an effect on which values are reported correctly. There are no differences observed for the median of the scores conditional on subjects being allocated to the Correct category between the HIGH and LOW treatment (MW: $z = 1.116$, $p = 0.2645$) or in the distribution of the values of the CorrectCalculations (Kolmogorov-Smirnov test: p = 0.527). This suggests that there is no particular impact on the behaviour of those reporting in a manner classified as Correct between the treatments.

There is a significant increase in the proportion of answer sheets assigned to the DeliberateError category between the HIGH treatment and the LOW treatment (FET: p = 0.022) and the CONTROL treatment (FET: p = 0.022).

The observation that there is no difference in the proportion of subjects whose response was categorised as Correct and no alteration in the distribution of the Correct responses between the LOW and HIGH treatments suggest that the major difference is in a switch of behaviour for a proportion of the subjects whose responses were categorised as a CalculationError in the LOW treatment to outcomes that were categorised as DeliberateCheating in the HIGH treatment. The higher propensity to report a PartialFalse value in the LOW treatment for intermediate values of CorrectCalculations compared to the HIGH treatment illustrated in Figure 2.11 relates to the greater level of answer sheets classified as CalculationError in the LOW treatment. This change in the pattern of reports suggests that the errors being made are not simply random and that the classification of CalculationError is an over estimate of the number of subjects making errors as it masks some subjects who were cheating at low levels, particularly in writing incorrect answers for harder problems or as the time pressure of the test came to bear. However as subjects were able to attempt the questions in any order they choose, and some clearly started from the bottom rather than the top, it would require a further level of subjective opinion as to which order a subject attempted the questions to gauge any such effect. Further research should include a control for the estimation of error rates, such as the removal of an incentive to cheat by having the experimenter mark and pay the subjects for the test.

Figure 2.12b shows the proportions of subjects' answers sheets that were assigned as DeliberateError for each of the minor categories identified as the manner by which subjects cheated. Interestingly while cheating was markedly increased in the HIGH treatment, the proportions of subjects among those cheating that were categorised to be deliberately cheating though any one of the three methods identified remained the same between the HIGH and LOW treatments. This suggests that alteration of the reference point did not serve as a trigger for a particular form of cheating, but merely to increase the incidence of each of the various mechanisms identified.

One particular behaviour that was observed on the answer sheets was subjects who cheated by reporting the maximum value while making no serious attempt at the numeracy test given to them in the experiment, other than to fill in the answer sheet with mostly random numbers or an incremental pattern. The proportion of subjects engaged in such maximal DeliberateCheating, that is exhibiting a pattern of ReportedCalclulations equal to AttemptedCalculations equal to 20 with a value for CorrectCalculations of 5 or fewer, is consistent over the treatments at 6.8%, 8.5% and 9.1% in the CONTROL, LOW and HIGH treatments. The similarity of these proportions between treatments indicates that the individuals who were prepared to act in this way were not affected by the reference point, consistent with agents in the model who have little or no cost of lying from either of the terms in the utility function. Subjects who reported in this manner are responsible for the high intercepts and negative slopes observed for the regression results of the NW quadrant in Figure 2.7.

2.8.7 The Effect of Gender

Dreber and Johannesson (2008) report that men were significantly more likely to lie in a sender-receiver game, though other authors have failed to detect such effects in similar experiments (Childs 2012; Gylfason et al. 2013). While the model presented here makes no prediction in relation to gender, subjects were asked to state their gender along with a number of other personal characteristics in the first part of the experiment in order to be used as controls.

The censored regression of the value reported by subjects shows a strong positive effect from being male ($p = 0.007$) when only gender is added (Model 2), though while a positive effect remains after the other terms of Model 3 are added, the observed effect is weaker ($p = 0.055$). There is a strong positive effect on the propensity to report a FullFalse value associated with being male ($p = 0.016$). Males reported a value classified as FullFalse value significantly more than females when considered across all treatments (FET: $p = 0.001$), with this difference most pronounced in the CONTROL treatment (FET: $p = 0.002$), though less so in the LOW treatment (FET: $p = 0.078$) and not significantly in the HIGH treatment (FET: $p = 0.165$).

2.9 Discussion

The aim of the study presented in this chapter was to propose and examine a model as to how people lie. The theory proposed contains two psychological costs, the first being an intrinsic measure that increases with the size of the lie being told, the second being a measure that increases with deviation above a reference point, representing additional disutility from reporting a value in excess of a perceived average or norm that may relate to a preference for the appearance of honesty, a concern for the credibility of the value being reported a desire not to appear boastful. The experimental results provide evidence consistent with the predictions of the model, that as well as reporting partial false values, a portion of subjects were observed to lie more in the treatment with a HIGH reference point compared to a LOW one.

The intuition of the key result is simple, the higher value of the reference point in the HIGH treatment served to make lying more acceptable for some subjects than in the case where the lower value was used. There are a number of potential reasons as to why the alternative reference point of the HIGH treatment may have changed subjects' behaviour. The statement that one of the previous most common reported values in previous sessions was twenty may have lead subjects to believe that a high proportion of subjects had cheated in the previous sessions, potentially indicating that cheating was possible, that cheating had occurred at a significant level or a combination of both in that cheating was possible and had occurred at a significant level. The information may therefore have served to reduce some subject's cost of lying by indicating the presence of the opportunity to cheat and (or) an existing norm of cheating in the experiment. This change in behaviour is inconsistent with a model for lying aversion containing only a cost in terms of the size of the lie.

The theory allows for a number of different types that exhibit varied behaviour in response to a change in the underlying state or a change in the reference point. As noted in the hypotheses section it is not possible in the experimental design used here to test for the type of a subject, but it is still possible to examine if the observed behaviour is consistent with the types suggested by the model.

A dishonest agent who has little or no cost of lying is expected to always report the maximum and this will not vary with treatment. The results find a common proportion of subjects in each of the treatments who make little effort in the experiment and always report the maximum value suggestive of such dishonest agents. An honest agent who has a high cost in terms of the size of the lie and a relatively low cost in terms of deviation from the reference point should report correctly in all treatments. The common rate of correct reporting observed for the three treatments suggests that there is some proportion of subjects who conform to this behaviour.

The model predicts that moderate agents with a low cost in terms of the size of the lie (such that they would be prepared to tell a lie) but a relatively high cost in terms of deviation from the reference point should report a lie that will decline in its extent as the realised state increases. Furthermore, such an agent will increase the values they report (for a given state) in the HIGH treatment compared to the LOW. The observed increase in the proportion of subjects reporting the maximum value in the HIGH treatment compared to the LOW treatment is consistent with this prediction. Furthermore the higher proportion of subjects found to be reporting a PartialFalse value in the LOW treatment compared to the HIGH treatment for intermediate values of the number of CorrectCalculations is indicative of such a moderate type. Indeed it is the change in behaviour of subjects of this type between the two treatments that gives rise to main result.

Interestingly the result in this chapter is different to that of Mazar et al. (2008) who found no difference when setting different reference points in an experiment based on self-reporting values for completion of a matrix test. It may be that their choices of previous average scores (four and eight) were insufficiently different to trigger an observable effect given that they could not observe individual cheating and found a very low proportion of subjects reporting the maximum value.

A key difference to other studies of the work presented in this chapter is the endogenous attribution of the realised states through the use of a numeracy test rather than through an exogenous lottery. The primary driver for this design feature was to make the subjects' decision relate to a self-generated value rather than one randomly assigned to them. A secondary feature is that the experiment is able to observe cheating at an individual level. The ability to record individual responses lead to the observation of a rich pattern in the different manners in which subjects cheated, from the deliberate and obvious, putting the same value for all twenty answers, to the much more subtle, putting correct answers to the first ten questions but then random numbers for a further five. While there is evidence that people cheat more and in different ways when the realised state is unobserved by the experimenter (Abeler et al. 2016; Hao and Houser 2017), the results presented here give further insight into some of the psychological mechanisms of cheating. The experimental design represents a key innovation of the chapter, in that future studies may use it to further examine dishonesty under the condition where the intention of the response is ambiguous.

The additional cost in terms of the deviation from the reference point represents the primary innovation of the model offered in this chapter. In a recent work, Abeler et al. (2016) proposed a number of models and concluded that one with both a cost in terms of the size of the lie and for a preference for appearing honest was consistent with experimental results. Such a model is also consistent with the experimental results presented here. The major difference is that the model detailed in this chapter does not explicitly model a preference for appearing honest, but rather includes a psychological term to reflect how a social norm or an agreed mode of behaviour may serve as an additional cost affecting an individual's decision. As described in the introduction, such considerations may be thought of as a preference for appearing honest or as a regard for the credibility of the value being reported. The chapter does not examine the causes of the cost in terms of deviation from the reference point, but simply investigates if such a model is consistent with experimental results. The chapter therefore does not attempt to address the different interpretations of the reference point in relation to a preference for appearing honest, the credibility of the reported value or a disutilty from boasting. Future work should therefore further examine the nature of the roles of a reputation for being seen to be honest and of the credibility of a lie in people's decisions when telling lies.

Chapter 3

An Experimental Examination of Lying in Different Sample Subject Pools

3.1 Introduction

A number of experiments have been conducted which demonstrate that the extent of a subject's honesty varies according to the context. Variation in the relative payoff to a subject with an opportunity to lie to that of the person affected has been observed to cause different proportions of subjects to report falsely (Erat and Gneezy 2012). Similarly, a greater propensity to lie has been observed under conditions where the intension of the act of lying could be viewed as less transparent (Hao and Houser 2017; Jiang 2013). The degree of social contact between subjects in an experiment has also been witnessed to be an important factor in the decision to report honestly, whereby some subjects were less likely to lie to friends compared to strangers (Chakravarty et al. 2011).

The experiments conducted in each of the studies cited in the previous paragraph involved a sample consisting of students attending a laboratory. Economics experiments are often conducted in a laboratory using student subjects as this method gives the researcher control over the environment and a convenient subject pool (Frechette 2016). However the use of students as subjects can be restrictive and a number of arguments have been made for extra-laboratory methods including the use of a more representative sample to increase external validity and to address issues of subject self-selection into experiments (Charness et al. 2013; Harrison and List 2004).

Recently, researchers have examined the potential role for burgeoning on-line labour markets such as Amazon's Mechanical Turk (AMT) service to serve as a source of subjects for economic experiments (Charness et al. 2013; Horton et al. 2011; Paolacci et al. 2010). There is evidence to suggest that such labour services mitigate two of the largest barriers to the widespread use of on-line experiments, the requirement to be able recruit and securely pay subjects and to ensure the internal validity of the experiment. The use of such services may be of great benefit to the field of experimental economics as they offer experimentalists the opportunity to recruit a low cost alternative to students that is potentially far more convenient and representative (Paolacci et al. 2010).

The use of different subject pools to examine behaviour has a long history in experimental economics (Frechette 2016). Many behaviours have been found to replicate between pools and some others can be replicated by exposure to the appropriate market. Confidence in the transferability of outcomes is even greater when considering treatment effects rather than point estimates (Frechette 2016). The availability of potential subjects through on-line labour services such as AMT has further driven studies to compare subject pools (Bartneck et al. 2015; Exadaktylos et al. 2013; Horton et al. 2011; Paolacci et al. 2010; Peer et al. 2014). Initial evidence suggests that there is little difference in responses between students and workers recruited from AMT in classic studies such as framing effects and simple games (Horton et al. 2011; Paolacci et al. 2010), though there is some evidence for a requirement to filter for subject understanding of the experiment (Horton et al. 2011).

The experimental context may well, however, be somewhat different for a worker through an on-line labour provision service rather than a typical student and, as described in the first paragraph, context can be an important factor in the decision to lie. One key difference may arise through the role of the worker's reputation, in the sense of the percentage of employer approved tasks previously undertaken by a worker. The rate of previously approved tasks is available as a filter to recruiters on the AMT system, such that unapproved work may serve to damage a workers reputation in that it may reduce their eligibility for future tasks (Kees et al. 2017; Peer et al. 2014). Indeed, Peer et al. (2014) present results demonstrating that high reputation workers produce higher quality data, proposing that researchers should use a cut-off based on the task approval rate for recruitment. A desire to maintain the value of their reputation may therefore affect AMT workers' decision making processes in an experiment differently to students. While there is little evidence for differences between subject pools and AMT workers in the outcomes for standard economic experiments, there is an open question as to whether such concerns for reputation may have an effect in a context where a subject has to make a moral decision such as whether or not to lie. It may be that a moral task constitutes a different form of decision for subjects recruited from a labour service than from a pre-registered pool of potential experimental participants.

The sender-receiver game forms the basis of one of the main mechanisms by which lying aversion has been examined (Erat and Gneezy 2012; Gneezy 2005; Gneezy et al. 2013). The experimental design involves the revelation of some state from among a set of states to one of two players, the sender, who must then send a message to the second player, the receiver, about the state. The message transmitted may or may not be truthful and the payoffs to each of the players may vary with the content of the message or some action chosen by the receiver after receiving the sender's message. A key design element of sender-receiver games is that while the party affected by the lie (the receiver) need not know whether the party with the potential to lie (the sender) has stated the truth or not ex post, the experimenter does observe the true state and the message of the sender. This feature contrasts to an alternative class of designs for the examination of lying where the experimenter does not observe an individual's actions, but compares sample outcomes to theoretical predictions or alternative treatments (Fischbacher and Föllmi-Heusi 2013; Mazar et al. 2008). Experiments with a design based on the sender-receiver game have an advantage in that the experimenter can observe and analyse individual lies. There is some evidence, however, that this advantage to the experimenter is offset to some degree by an effect on behaviour arising from the design, as evidenced by differences in behaviour detected for identical lotteries between subjects whose decisions were observed to those who were not (Abeler et al. 2016; Gneezy et al. 2016).

This chapter describes an on-line implementation of a sender-receiver game undertaken by separate subject pools of undergraduate students and of workers recruited through the AMT system. This study is not the first to perform a sender-receiver game with AMT subjects, as Biziou-van-Pol et al. (2015) conduct such a task as part of their investigation into the relationship between lying behaviour and pro-social preferences. The aims of the experiment presented in this chapter are, however, somewhat different. The first aim of the investigation is to examine if there are substantial differences between the reporting behaviour observed for the two samples. This chapter is the first to present results detailing a comparison of different subject pools undertaking a sender-receiver game. The second objective of the chapter seeks to examine if subject responses are altered by a change in the distribution of states underlying the sender-receiver game through a treatment that places a higher probability on lower values. This second objective parallels treatments recently described in Abeler et al. (2016) and Gneezy et al. (2016), who detail self-reports of unobserved lottery draws by subjects with different distributions of payoff values. The difference in the approach used in this chapter is that the draw presented to the subjects is observed. A further key difference to previous studies using the sender-receiver game to investigate deception (Biziou-van-Pol et al. 2015; Erat and Gneezy 2012; Gneezy 2005; Gneezy et al. 2013) is an expansion of the size of the set of states used from 2 or 6 to 100, an innovation also investigated in Gneezy et al. (2016).

The chapter makes two methodological contributions. The first contribution is to quantify differences in behaviour between a standard undergraduate subject pool and and an AMT worker pool in a simple online deception game. The second contribution pertains to the extent the subjects engage with and understand the task, particularly the experimental instructions, and how that leads to decisions that be attributed to confusion, as opposed to non-standard preferences (Akiaya et al. 2017; Andreoni 1995; Houser and Kurzban 2002).

3.2 Literature

In many economics experiments, subjects are invited to a session in a laboratory from a pool of people who have pre-registered to take part in experiments, often using the system ORSEE (Greiner 2015). These subject pools typically consist of students from the university of the researcher(s). However two major concerns have been raised in relation to the use of pools of students as subjects for experiments (Harrison and List 2004). The first issue is that such pools consist of people who have volunteered to undertake experiments, such that there is some degree of self selection. The second concern is that the pool consists of students who may not be representative of a wider population. Frechette (2016) surveys a number of studies conducted with subject pools that differ from that typically used comprising of student subjects including animals, children, representative samples and professionals. The author draws a general conclusion that results carry over between different groups with a number of exceptions to this rule.

A number of papers address the issue of self-selection of subjects by volunteering to take part in experiments (Anderson et al. 2013; Cleave et al. 2013; Falk et al. 2013; Slonim et al. 2013). There are potentially two competing reasons as to why such self-selection may form an issue (Falk et al. 2013). Potential subjects may be driven by the payments on offer and therefore be more self-serving than the more general population. Alternatively, the possible participants may have a desire to assist with the research being conducted and therefore be more pro-social than the general population.

Falk et al. (2013) compare the charitable donations of students made before arrival at a university with their decisions to undertake experiments in the laboratory. The authors report no significant relationship between the student's charitable giving and their participation in experiments. Cleave et al. (2013) examine the risk and social preferences of a first year undergraduate class among those who subsequently attend laboratory experiments and those who do not. The results show no relationship between social preferences or risk attitude and attendance at experimental sessions, though a significant relationship between the amount sent in a trust game and a lower propensity to attend experiments suggests that the extent of pro-social behaviour in the lab may be less than in the population. Slonim et al. (2013) report a similar experiment recording a number of survey responses as well as experimental decisions and report that participants who had lower income, more leisure time, more interest in economics and were more pro-social on the dimension of volunteering time were more likely to participate in experiments. Anderson et al. (2013) further report a similar result in a comparison of pro-social preferences between a typical sample of university students and trainee truckers recruited within a residential training program.

Further papers, or other sections of some of the papers detailed above, seek to examine the question as to whether typical student subjects exhibit the same behaviours as a wider, more representative, sample over a number of different economic decisions. A primary category of interest has been social preferences. Bellemare and Kröger (2007) examine social capital, as measured by choices in a trust game, and report that the student sample serves to form a lower bound on the degree of pro-social preferences, a result that is also reported in Falk et al. (2013). Differences in pro-social behaviour between a community group and students, as recorded in a representative dictator game with a charitable donation, are also presented by Carpenter et al. (2008), where members of the community group are reported to be significantly more pro-social than the students. Cappelen et al. (2015) detail similar findings using a dictator game and a trust game between students and a representative sample of the Norwegian population and the same pattern of lower pro-social behaviour is also reported between a sample of self-selected students and one of self-selected adults in Anderson et al. (2013).

Other studies have examined for differences in risk attitude and inter-temporal preferences (Andersen et al. 2010; Gaudecker et al. 2012), in co-ordination through public goods games (Bortolotti et al. 2015; Gächter et al. 2004) or the prisoner's dilemma (Bigoni et al. 2013), in depth of reasoning in beauty contests (Bosch-Domènech et al. 2002) and for differences in bidding behaviour in auctions (Depositario et al. 2009). Belot et al. (2015) provide evidence of consistent outcomes for a similar set of games with inexperienced samples of students and non-students and conclude that results based on student samples are likely to over-estimate the extent of selfish and rational behaviour in the wider population.

An alternative set of investigations, focusing on whether the choices of students differ from those with professional experience, particularly in a domain relevant to the choice under examination, are reviewed in Frechette (2015). Experience of a particular domain, rather than expertise, may also have an effect on choice. One area of research where students typically have little experience compared to the wider population is in tax filing, such that researchers have investigated the differences between student and non-student sample pools is tax compliance (Alm et al. 2015; Choo et al. 2016). Alm et al. (2015) report no significant differences between student and non-student subject pools other than where specific external experience of a treatment feature may have a role. Choo et al. (2016) on the other hand report higher levels of compliance in taxpayer samples compared to a sample of students which they attribute to the application of norms of compliance from outside the laboratory. The lower degree of non-compliance observed among taxpayers compared to students in Choo et al. (2016) reveals that the context may have different effects on different samples when making moral choices.

In a paper examining lying, Abeler et al. (2014) compare responses in a 4 coin flip task where subjects were paid 5 EUR for each tail flip reported over the telephone between members of the German public at home and students from a German university in the laboratory. Students in the laboratory reported significantly more tails then the members of the public. Results among the student sample were similar in a further treatment where they were required to submit their value through a computer. Ruffle and Tobol (2017) report on lying behaviour among a sample more representative of the general population that students using current and former members of the Israel Defence Force in a self-reported dice roll task. The authors detail a correlation between a lower tendency to over-report the value of the die with higher scores on tests of cognitive ability.

Two key elements of the literature discussed in the previous paragraphs are the mechanism by which the non-standard sample was recruited and way in which the experiment was conducted. Among the papers cited, a number of different methods for recruitment were utilised, including contact with a suitable employer (Burks et al. 2009), attachment to established surveys (Bellemare et al. 2008), the use of market research firms (Choo et al. 2016), extending laboratory sign ups to non-students at a campus (Alm et al. 2015), newspaper articles (Bosch-Domènech et al. 2002) and the use of e-mail and advertisements (Belot et al. 2015). Methods for conducing the experiment included calling subjects into the laboratory, establishing local laboratories in hotels (Andersen et al. 2010), telephone calls (Abeler et al. 2014), utilising sessions on training courses (Burks et al. 2009) and conducting experiments on-line (Choo et al. 2016). The time and financial cost of such methods, typically allied to the higher experimental cost of incentive payments due to the higher opportunity costs of non-student subjects, from part of the rational for why student subjects are typically used.

Recent technological advances have, however, produced on-line labour markets that offer a low cost, highly scalable and potentially more representative sample to researchers (Horton et al. 2011). Services such as AMT offer the advantage of addressing a number of difficulties and concerns around using on-line experiments, in particular in relation to the recruitment of subjects, making secure payments to them and ensuring the internal validity of the experiment.

Paolacci et al. (2010) detail an experiment comparing survey responses to three common tasks from the heuristics and biases literature among samples of student subjects recruited through a laboratory, AMT workers and subjects recruited through internet message boards. The results show that responses did not differ significantly between the pools and errors in responses were lower for subjects recited through AMT then through internet discussion boards. Demographic data suggests that AMT workers are more representative of the U.S. population than student sample pools (Paolacci et al. 2010). Horton et al. (2011) report results for a number of investigations conducted using AMT workers, notably that priming and framing produce effects consistent with findings from the laboratory. In addition the authors compare choices from a prisoner's dilemma conducted with AMT workers to those of students in a laboratory and find the selections do not differ, especially when subjects who failed a comprehension test are excluded.

The over-communication of truthful messages compared to equilibrium predictions has been observed in a number of experimental studies of the sender-receiver game (Cai and Wang 2006; Sánchez-Páges and Vorsatz 2007; Wang et al. 2010). An aversion to lying was observed through the different propensity for choices over the same pairs of payoffs between a sender-receiver game a dictator game where the higher payoff to the sender required sending a message that was untrue to the receiver (Gneezy 2005). Sánchez-Páges and Vorsatz (2009) demonstrate that lying aversion is due to a desire not to lie rather than a preference for telling the truth. Evidence for pure lie aversion, rather than arising from beliefs is presented by López-Pérez and Spiegelman (2013). Erat and Gneezy (2012) demonstrate that the propensity to report a false value is a function both of the sender's own and the receiver's payoff, and produce a classification of lies in terms of the benefit or harm both to the sender and the receiver. Gneezy et al. (2013) describe results that highlight different types of subject who report honestly for some values of realised state, but report falsely for others.

Biziou-van-Pol et al. (2015) report on an examination of a sample of workers recruited through AMT into the relationships between altruistic preferences and lying aversion and between co-operative tendencies and lying aversion. The authors report a difference in the proportions reporting a false value between treatments of a sender-receiver game with a Pareto improving set of payoffs and an altruistic pair of payments and that men are more likely than women to tell an altruistic white lie, but not to tell a Pareto white lie. The results also demonstrate a positive correlation between honesty in the Pareto improving treatment and altruism in a dictator game that was not observed for the altruistic deception game. The authors also report a positive correlation between honesty in the Pareto improving deception game and co-operation in the prisoner's dilemma that was not found for the altruistic deception game.

Recently, both Abeler et al. (2016) and Gneezy et al. (2016) have reported on experiments where subjects were required to report values based on draws from different distributions. Abeler et al. (2016) report on results demonstrating "drawing-in", whereby the proportion of subjects reporting a true low value compared to the theoretical value decreases as the probability of the low state decreases. Likewise, Gneezy et al. (2016) report on results showing that subjects demonstrate a higher propensity to report a partial lie with a decrease in the probability of the highest payoff state.

The design of the sender-receiver stage of the experiment of Biziou-van-Pol et al. (2015) reflects the design used in Erat and Gneezy (2012) whereby there are two potential states and a sender subject must report a state to the receiver subject and separate treatments that vary the relative payoff of the two states to the sender and receiver which are unknown to the receiver. The design differs in that the message is used to determine the payoff to both players, rather than a subsequent action chosen by the receiver, an alteration similar to that used in Gneezy et al. (2013), whereby the message transmitted determined the sender's payoff, though the payoff to the receiver was a function of the receiver's action. The experimental design used in this chapter further extends the number of potential states to 100 while retaining the feature that the receiver is passive and that only the message sent by the sender will affect the payment to the two players.

3.3 Experimental Design and Procedures

The experiment was conduced with two separate samples, a first consisting of undergraduate students at the University of Exeter and a second of people registered as workers in the Amazon Mechanical Turk (AMT) system. This section will first describe the design of the sender-receiver task before detailing other common elements of the experiment and the recruitment processes used for the two samples.

3.3.1 Experimental Design

The experiment was conducted on-line using bespoke software running on a web server hosted by the University of Exeter. Full details of the experimental instructions are given in the Appendix.

The third part of the experiment was a sender-receiver game framed as a lottery draw over values in the range 1 to 100. Figure 3.1 shows the instructions for the baseline treatment. Subjects were informed that they would be randomly and anonymously paired with another subject from the experimental session, and that their paired subject would receive the same instructions and their own independent draw from the lottery. The instructions informed subjects that there were two roles in the experiment, a type A role (corresponding to a sender) and a type B role (receiver), and that one member of each pair would be assigned to one of the roles but that the actual role they were to be assigned to would be determined after the experiment was completed. The instructions detailed that each of the pair of subjects would be paid the same multiple of the value reported by the person allocated to be of type A and their task was to report the value they would wish to report if assigned to the type A role.

Two treatments in which the distribution underlying the lottery was varied were performed. Figure 3.1 shows the instructions for the baseline condition, BASE, where a uniform distribution with each of the values in the range 1-100 being equally likely was used. In the treatment condition, SKEW, the lottery values

FEELE Laboratory

Experiment Part 3 - Instructions

This is the final part of the experiment. In this part you have the opportunity to increase your payment from today's experiment. The payment for this part of the experiment will be in addition to your base payment, as detailed earlier in the experiment.

The computer will perform a lottery by drawing a random whole number from 1 to 100, with all numbers having the same probability of being selected. That is the chance of the value 12 being drawn is 1-in-100, as is the chance of the value 77 being drawn.

An alternative way to think about the lottery would be that the computer is to make a draw of one ball from an urn containing 100 balls. Each of the balls in the urn has a different number on it, from 1 to 100.

There are two types of players in this part of the experiment: Type A and Type B.

Type A roles report the value of their lottery draw to a Type B player chosen at random.

You will not know what type you will be allocated before you make your choices. We will ask you to make the choice as to what message you would send to the other person in your pair if you were of type A.

Once the experimental session is over, half of the participants in part 3 of the experiment will be randomly selected by the computer to be of type A. The remaining participants will be assigned to the role of type B. The computer will then randomly and anonymously match each participant selected to be of type A with one assigned to be type B.

The computer will calculate your payment for part 3 of the experiment based on the type you are selected to be and the choices made. Payments will be calculated in the following manner:

For the person selected to be **type A**, the payment will be 5 cents times the value they report

For the person selected to be **type B**, the payment will be 5 cents times the value the type A person they have been paired with reports

The following table presents 4 examples of outcomes. In each case the lottery produces the same values for the two players (you and your paired other). The differences in the payments arise from the values chosen and the random selection of who is assigned to type A and who is assigned to type B.

	Example 1		Example 2		Example 3		Example 4	
	You	Other	You	Other	You	Other	You	Other
Value Drawn	30	60	30	60	30	60	30	60
Value reported	30	60	30	60	80	20	80	20
Type Assigned	A	B	B	A	A	B	B	A
Payment	$1.50	$1.50	$3.00	$3.00	$4.00	$4.00	$1.00	$1.00

Only the value you choose to report for the type you are selected to be will be shown to the experimental participant you are paired with. The actual value drawn in the lottery will not be revealed to the other participant.

Your assigned type and the value of your payment will be available online after you have completed the experiment. There may be a delay before the final result is available while we wait for other subjects in order for you to be matched to someone.

Next

Please do not use the back button on your browser during the experiment
Logout

Figure 3.1: Screenshot of instructions for BASE treatment

in the range 1-37 had a higher likelihood of being drawn (3-in-174) than higher values (1-in-174). This distribution was chosen as the maximum value of the range given the higher probability corresponds approximately to the expected value of the distribution. Table 3.1 summarises the treatments and sample sizes.

Sample	Treatment	Sample	Male
Student	BASE	50	48.0%
	SKEW	51	47.1%
AMT	BASE	179	61.5%
	SKEW	181	61.9%

Table 3.1: Summary of treatments and sample sizes

3.3.2 Experimental Procedures

Upon access to the software, subjects were informed that the experiment involved three parts and that they would be required to complete the second part of the experiment in order to qualify for the third part of the experiment. They were further informed that the first part of the experiment would involve a number of questions for which they were required to make selections. On the first page of the first part of the experiment, subjects were presented with a questionnaire eliciting personal details such as age and gender as well responses to questions in the relation to the five factor model of personality traits (Woods and Hampson 2005). The subsequent pages of the first part of the experiment asked subjects to undertake a hypothetical risk-aversion task based on Gneezy and Potters (1997) and a hypothetical inter-temporal investment decision.

The second part of the experiment informed subjects that they were required to undertake a slider repositioning task as described in Gill and Prowse (2012). The instructions informed subjects that there would be three practice rounds before a final qualification round in which they would be required to reposition at least 15 sliders to the centre of the bar in 60 seconds in order to qualify for the third part of the experiment.

Subjects who failed to complete 15 or more sliders in the final round of the slider task were informed that they had failed to qualify and would receive only their show up fee. Subjects who repositioned 15 or more sliders were shown instructions for the third part of the experiment.

Before progressing to the lottery draw and making their decision, subjects were given a short multiple choice quiz on the experiment to test their understanding of the lottery reporting task. The six questions used are given in the Appendix. Incorrect answers were reported and further information given to the subjects. Correct answers to all questions were required in order for subjects to progress to the reporting task. Once the quiz was completed, subjects were informed of their lottery draw and asked to submit the value they would like to report.

3.3.3 Student Sample

Subjects were recruited from a pool of previously signed up University of Exeter students subjects using ORSEE (Greiner 2015). Students studying psychology or in the final year of an economics course were excluded from the recruitment process. Subjects in the student sample were informed as part of the ORSEE recruitment process that the experiment would be conducted on-line, that they would be able to access the experiment over a three hour period in a pre-defined window, that the experiment should take no longer than an hour to complete, payment for the session would be made to a registered bank account and there would be no need to attend the laboratory. Student subjects who signed up for the session were informed that they would receive further details of how to access the experiment prior to the session starting.

Sessions were conducted between 9 am and 12 am. Subjects who had registered to undertake the experiment were sent an e-mail the evening before their session containing a unique username and password as well the address of the appropriate web site for the experiment. The e-mail reminded them of the time window and that payments would be made to their bank accounts.

Upon accessing the software, subjects were prompted to enter their username and password. The experiment then progressed as described in the previous section with a number of specific additional features. Student subjects were informed as part of the initial instructions that the show up fee for the session was £3. For the third part of the experiment, Student subjects were informed that the person selected to be type A would be paid 5 pence times the value they reported and the person assigned to be type B would be paid 5 pence times the person selected to be type A reported. At the end of the session student subjects were given further information about their payment and how it would be made. This required that they enter their bank details into the university's payments system and confirm through the software that this had been carried out. Subjects were also e-mailed this information once the session was completed. In a small number of cases, subjects had to be further contacted to confirm their payment details.

Sessions were run between 12 Jan 2016 and 5 Feb 2016. Two sessions each recruiting 40 subjects were run for the BASE treatment, and a third session recruiting 80 subjects was run for the SKEW treatment. The median payment to student subjects was £6.75 for a task with a median time of 18 minutes and 47 seconds, corresponding to a median payment rate of £21.56 per hour, with a lowest value of £4.60 per hour and a highest value £36.20 per hour.

3.3.4 AMT Worker Sample

AMT subjects were recruited through advertising a HIT on the Amazon Mechanical Turk system. Subject recruitment was pre-filtered in the AMT system according to several criteria for the potential workers, that they must have a number of previous HITs approved greater than 500, a HIT approval rate greater than 95%, be located in the United States and not have been assigned a role created by the experimenters to identify subjects who had undertaken the experiment. The recruitment message shown in AMT is given in the Appendix. AMT subjects were informed that the fee they would receive for the experiment was $3 and that they could earn more through their choices in the experiment that would be paid in the form of an AMT bonus. Upon reading the recruitment message on AMT, subjects could accept the HIT, whereby they would have an hour to complete the experiment and validate their payment. If a subject failed to register their completion, in particular because they failed to complete the experiment in the time frame or chose not to do the experiment upon reading the instructions, the HIT would be republished for another worker to access.

Upon accessing the software, AMT subjects were shown details of the ethical considerations of the experiment and a set of contact details for the experimenters (given in Appendix). Subjects who progressed were invited to enter the web site using their unique AMT identifier. The remainder of the experiment progressed in the same manner as given before with a number of minor differences. The instructions for the first part of the experiment detailed a show-up fee of $3, task values were presented in dollars rather than pounds, and the payments in the third part of the experiment were detailed as 5 cents times the value reported by the person assigned to the type A role. At the end of the experiment, AMT subjects were given a unique code generated by the experimental software that they were required to enter into AMT such that the match between their AMT identifier and their software code prior coud be verified to any payments being actioned.

Sessions were run between 10 March and 21 April 2016, with subjects recruited on specific days in batches of 60. The number of subjects assigned to each treatment is given in Table 3.2. AMT subjects were assigned sequentially to one of the treatments upon login to the software. At the end of each session, subjects were paid their fees and bonuses and allocated a label in the AMT system such that the subjects who had undertaken the experiment could be identified and excluded from future sessions.

The median payment to AMT subjects was $6.00 for a task with a median time of 16 minutes 45 seconds, corresponding to a median payment rate of $21.49 per hour, with a lowest value of $5.16 per hour and a highest rate of $45.15 per hour. The average AMT payment rate is considerably higher than the minimum wage in the US and higher than rates typically quoted as being paid for performing surveys on AMT, which have been seen to be as low as $1.38 (Horton and Chilton 2010).

3.3.5 Errors

The proportion of AMT subjects observed making at least one error in the quiz over the BASE and SKEW treatments (81.5%) is higher than that found for the student sample (73.8%) (Fisher's exact test (henceforth FET): $p = 0.07$). These proportions are, however, notably high for both samples. Subjects who gave incorrect answers on the quiz were prompted with further information and invited to answer the questions again. The proportion of AMT subjects making a repeated mistake in the quiz, conditional on making a mistake at the first attempt (10.9%), was found to be significantly lower than for the student sample (23.3%) (FET: $p = 0.005$). Subjects who incorrectly answered a question on the quiz more than once are excluded from the main analysis. A larger fraction of student subjects (17.2%) was excluded than AMT subjects (8.9%) (FET: $p = 0.012$).

3.3.6 Sample Characteristics Comparison

Males were slightly initially under-represented in the sample of student subjects, comprising 43.5% of student subjects recruited, whereas males were initially over-represented in the AMT sample (54.7%), atypical of previous studies (Paolacci et al. 2010). The proportion of females who failed to qualify for the final part of the experiment in the slider task or who were excluded for making a repeated error in the quiz is higher than for males in both samples, leading to a more balanced male to female ratio in the student sample (47.7%) but to a more exaggerated male to female ratio in the AMT sample (59.2%).

The median age of a subject recruited through AMT (32) is significantly higher than that of the student sample (19) (Mann-Whitney (henceforth MW) test: $z = 15.053$, $p < 0.001$), reflecting the different nature of the subject pools (Paolacci et al. 2010). A number of differences were found between the samples in the self-reported personality measures. Student subjects were more likely to identify themselves as extroverts compared to subjects recruited through AMT (MW test: $z = 4.860$, $p < 0.001$) and to have a lower degree of emotional stability (MW test: $z = 5.029$, $p < 0.001$). No difference was found in the median value allocated to the risky component of the investment task.

The median length of time taken by the AMT subjects over the combined BASE and SKEW treatments was shorter than that taken by the students (MW test: z = 4.410, $p < 0.001$). A difference in the median time taken between the samples was found for a number of sections of the experiment, and this finding is discussed in further detail in the results section.

3.3.7 Sample Summary

The numbers of subjects who were recruited and the number who completed the first part of the experiment including qualifying for the final bonus payment round are shown in Table 3.2. The table also details the number of subjects making errors in the quiz to ascertain understanding before undertaking the decision in part three of the experiment as well as the number of subjects used in the analysis and the number of the used subjects who reported being male.

Sample	Treatment	Recruited	Completed	Error	Repeated Error	Used	Male
Student	BASE	80	62	46	12	50	24
	SKEW	80	60	44	9	51	24
AMT	BASE	237	191	156	12	179	110
	SKEW	241	204	166	23	181	112

Table 3.2: Sample characteristics by treatment

3.4 Hypotheses

The goal of the research presented in this chapter is to examine if subjects recruited from AMT behave in the same manner as subjects from a typical student subject pool when faced with a moral choice in a sender-receiver game whereby reporting a false value will increase both their payment and the payment to the person to who they report the value (a Pareto White Lie, as according to the classification of Erat and Gneezy (2012)).

There are a number of reasons as to why the choices made by the subjects in the two pools may be different. A first simple reason is that the demographics of the two pools will be different, for example the student subjects will on average be younger than the AMT subjects (Paolacci et al. 2010). A second potential reason is that subjects recruited through AMT are termed as workers in the AMT system. There are two mechanisms available to make a payment to a worker recruited through AMT. The first mechanism is the default, which involves a base fee which is paid upon approval of a successful unit of work completed by the AMT worker. This payment can be see to parallel a show-up fee used in experiments for student subjects. The second, additional mechanism, allows the recruiter (experimenter) to make a further payment in the form of a bonus. Such an optional bonus payment can therefore be used to make payments relating to the incentivised element of an experiment. A key filter for the recruitment of workers through AMT is the HIT approval rate, the proportion of previously attempted units of work successfully completed by a worker. A potential recruiter may filter for workers who only have an approval rate above some required level, indeed it has been suggested that researchers should employ such a filter and report the value used when performing experiments with AMT subjects (Kees et al. 2017; Peer et al. 2014). As such gaining the successful approval of a task could be of importance to AMT workers as it may influence their ability to undertake future work. While the separation of the processes for the approval and base payment from any bonus payment allows for the choice in the experiment of the AMT worker to be independent from their reputation in terms of the approval rating, some AMT subjects may be concerned that an immoral choice in the experiment will have a negative influence their reputation. As student subjects are specifically recruited to undertake experiments, rather than to serve as workers, such a concern will not be present, leading to a first hypothesis.

Hypothesis 1 *AMT subjects will report positive false values less frequently than student subjects*

The experiment used in the study presented here differs from a number of previous studies (Biziou-van-Pol et al. 2015; Erat and Gneezy 2012; Gneezy 2005) in that there is an expanded set of possible states underlying the sender-receiver game, as reported in a number of recent studies (Abeler et al. 2016; Gneezy et al. 2016, 2013). A larger number of possible states allows for subjects to report partial lies, as reported by Fischbacher and Föllmi-Heusi (2013) and further discussed by Abeler et al. (2016), and for the experimenter to adjust the distribution of possible states, as demonstrated by Abeler et al. (2016) and Gneezy et al. (2016). A second hypothesis can therefore be derived from the theory proposed in Chapter 2. Chapter 2 described that under an increase in the reference point, the minimum value reported by an agent is non-decreasing. A treatment with a higher expected value, in that the distribution of possible states is uniform compared to an alternative that is skewed to have higher probabilities for lower states, can be considered to have a higher reference point.

Hypothesis 2 *Subjects should misreport no less frequently in a treatment with a uniform distribution compared to one with a distribution skewed to lower values*

3.5 Results

3.5.1 Variables

A number of variables are referenced in the following section. The variable SKEW is binary and takes a value of 1 for measures from the SKEW treatment and 0 otherwise. STUDENT is a binary variable that takes a value of 1 if a subject is part of the student pool and 0 if the subject was recruited from AMT. LotteryDraw refers to the lottery draw presented to the subject prior to reporting their choice of value. ReportedValue is the choice made by the subject as the value to report. The variable Error is a binary measure which takes the value 1 if a subject answered any of the questions on the quiz of understanding incorrectly. The binary variable Male takes a value of 1 if a subject reported being male and 0 otherwise. The term Age refers to the age reported by the subjects. The terms Extroversion, Agreeableness, Emotional Stability, Conscientiousness and Openness refer to answers given on a scale from 1 to 7 about the subjects degree of agreement with two statements assessing their personality for each of these traits (Woods and

Hampson 2005). The pairs of statements used for this assessment are given in the Appendix. The variable Allocation is the subject's choice for investment in the risky component in the risk aversion measure of the first part of the experiment. The term Slider represents the subject's score in the (final) qualifying round of the slider task, which takes a minimum of 15 and maximum of 48, representing the number of sliders required to qualify for the final part of the experiment and the total number of sliders.

There are four terms widely used that refer to the time taken by subjects in various steps of the experiment. The term Prelottery is the total elapsed time between in seconds a subject login and being presented with the instructions for the final lottery reporting task. The variable Lotteryinstructions is the length of time in seconds a subject spent between loading the instructions for the lottery task and progressing to the quiz to check for subject understanding of the task. The term Lotteryquiz is the length of time a subject spent on the quiz. The term Lotterydecision is the length of time in seconds between a subject loading the lottery reporting decision page after completing the quiz and submitting a choice.

3.5.2 Classification of Choice

The choices made by subjects can be classified into one of three categories. The first classification is an UpLie, whereby the value of ReportedValue is greater than that of the LotteryDraw. The second classification is an Honest value, whereby the ReportedValue is equal to the LotteryDraw. The third classification is a DownLie, where the subject's choice of ReportedValue is less than the LotteryDraw.

The choices of subjects are analysed using a multinomial logistic model (Nerlove and Press 1973). Let y_i denote the classification of the choice made by the subject i with some associated vector of observable characteristics $\mathbf{x_i}$. The probability that a subject chooses a particular classification, j, from the a set of classifications 0, 1 ... J is given by $P(y_i = j|\mathbf{x_i})$. The multinomial logit model for the subject's choice can be expressed as the following:

$$P(y_i = j|\boldsymbol{x}_\mathbf{i}) = \frac{exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_j})}{\sum_{k=0}^{J} exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_k})} \tag{3.1}$$

To remove the indeterminacy in the model, a normalisation was conducted for one of the subjects' choices (Honest) such that the values are zero ($\boldsymbol{\beta}_0 = 0$), and the expression for the model can be rewritten as:

$$P(y_i = j|\boldsymbol{x}_\mathbf{i}) = \frac{exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_j})}{1 + \sum_{k=1}^{J} exp(\boldsymbol{x}_\mathbf{i}'\boldsymbol{\beta_k})} \tag{3.2}$$

The basic expression for the independent variables in the regression is formed from the interactions of the sample (STUDENT), the treatment (SKEW) and the value of the LotteryDraw, such that for a given classification (j):

$$\begin{aligned} \mathrm{x}_\mathbf{i}'\boldsymbol{\beta}_\mathrm{j} = \beta_{\mathrm{j}0} + \beta_{\mathrm{j}1}\mathrm{LotteryDraw} + \beta_{\mathrm{j}2}\mathrm{STUDENT} + \beta_{\mathrm{j}3}\mathrm{SKEW} + \\ \beta_{\mathrm{j}4}\mathrm{LotteryDraw} * \mathrm{STUDENT} + \beta_{\mathrm{j}5}\mathrm{LotteryDraw} * \mathrm{SKEW} + \\ \beta_{\mathrm{j}6}\mathrm{STUDENT} * \mathrm{SKEW} + \beta_{\mathrm{j}7}\mathrm{LotteryDraw} * \mathrm{SKEW} * \mathrm{STUDENT} \end{aligned} \tag{3.3}$$

An extended model (Model 2) for the regression further incorporates other variables relating to the personal characteristics of the subjects and the time taken by subjects for various elements of the experiment. In the first instance the three classifications detailed at the start of this section (UpLie, Honest and DownLie) are used and the regressions are normalised such that an Honest response serves as the baseline. In a later section the classifications are further divided but and regression model extended to reflect the additional classifications.

Greene (2010) proposes a mechanism for addressing issues relating to the usage of hypotheses tests of the coefficients of interaction effects in non-linear models raised by Ai and Norton (2003). The approach to analysis suggested by Greene (2010) is to perform tests relating to specification of the model prior to analysis, for which the author argues that graphical presentations create an informative adjunct to any statistical analysis. I adopt this recommended approach in the analysis of the results. The results presented make extensive use of the

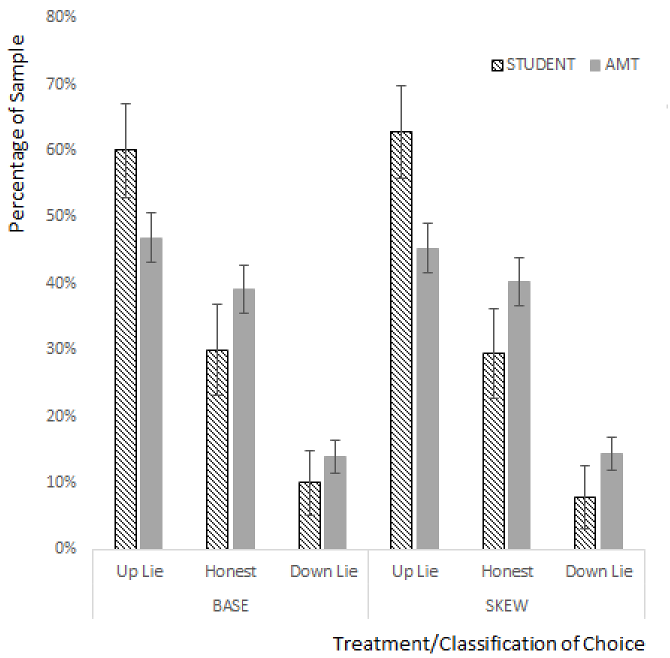


Figure 3.2: Proportion of subjects reporting by classification

post-estimation calculated propensities to report a particular classification for a given set of values of the LotteryDraw (0 to 100 in steps of 10). An overview of the approach used for the analysis is given in Williams (2012). Results for the regressions used are presented in the Appendix. All regressions use robust standard errors. Vertical or horizontal bars on figures represent standard errors.

3.5.3 UpLie

Result 1 *Student subjects demonstrated a higher propensity to report an UpLie than AMT subjects at values of the LotteryDraw below 50*

Figure 3.2 shows the proportion of responses by classification in the BASE and SKEW treatments for student and AMT subjects. The figure shows that proportion of subjects reporting an UpLie is larger among the student subjects than among AMT subjects in both treatments. This difference is significant for the SKEW treatment (FET: p = 0.039), though not for the BASE treatment (FET: p = 0.112). There are no differences in the proportions reporting an UpLie between the SKEW and BASE treatments in either the student sample (FET: p = 0.839) or the AMT sample (FET: p = 0.833).

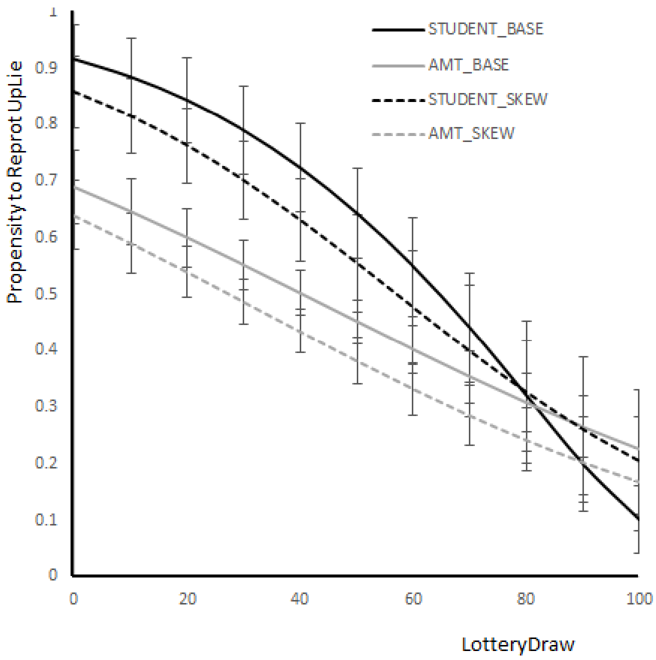


Figure 3.3: Propensity of subjects to report a value classified as an UpLie

There are two main factors for the different levels of the significance observed for the difference in the proportions of subjects reporting an UpLie between samples in the two treatments. The first relates to the smaller size of the student sample leading to relatively larger standard errors. The second factor relates to the range of vales of the LotteryDraw for which subjects were observed to show a higher propensity to report an UpLie. Figure 3.3 illustrates the propensity to report an UpLie as calculated at specific values of the LotteryDraw for the two samples and the two treatments.

The propensity to report an UpLie declines with the value of the LotteryDraw for both samples over both treatments. At high values of the LotteryDraw the propensity to report an UpLie is low and the same in the two samples. However, at low values of the LotteryDraw, the propensity to report an UpLie is greater among the student subjects then for AMT subjects. Values for the test of the contrast of the propensity to report an UpLie are given in Table C.2 in the Appendix. There is a significant difference in the calculated propensity to report an UpLie for values of the LotteryDraw of 50 or less in both the BASE treatment ($p \leq 0.0317$) and the SKEW treatment ($p \leq 0.0697$). The difference in significance of the non-parametric tests of the propensity to report and UpLie between the samples observed between the BASE and SKEW treatments presented in the previous paragraph can therefore also be related to the higher propensity for an UpLie occurring for lower values of the LotteryDraw combined with the greater manifestation of low values of the LotteryDraw in the SKEW treatment. These results are consistent with the first hypothesis.

3.5.4 Honest

Result 2 *AMT subjects show a higher propensity to report an Honest value for values of the LotteryDraw less than 20*

Figure 3.2 indicates a higher proportion of AMT subjects report an Honest value compared to the student subjects in both the BASE and SKEW treatments. Non-parametric tests, however, reveal no significant differences in the propensity to report an Honest value between subjects in the BASE treatment (FET: p = 0.252) or the SKEW treatment (FET: p = 0.192). There are also no differences between the treatments in the student sample (FET: p = 1.000) and the AMT subjects sample (FET: p = 0.830).

Figure 3.4 illustrates the calculated propensity to report an Honest value at a variety of values of the LotteryDraw for the student and AMT samples in the BASE and SKEW treatments. Values for the test of the contrast of the propensity to report an Honest value are given in Table C.3 in the Appendix. The calculated propensity to report an Honest value in the BASE treatment is higher among AMT than student subjects for values of the LotteryDraw of 20 or below ($p \leq 0.0676$) and for values of the LotteryDraw of 40 or below in the SKEW treatment ($p \leq 0.0926$).

3.5.5 DownLie

Figure 3.2 shows little difference in the proportion of student subjects reporting a DownLie compared to AMT subjects in the BASE treatment (FET: p = 0.636)

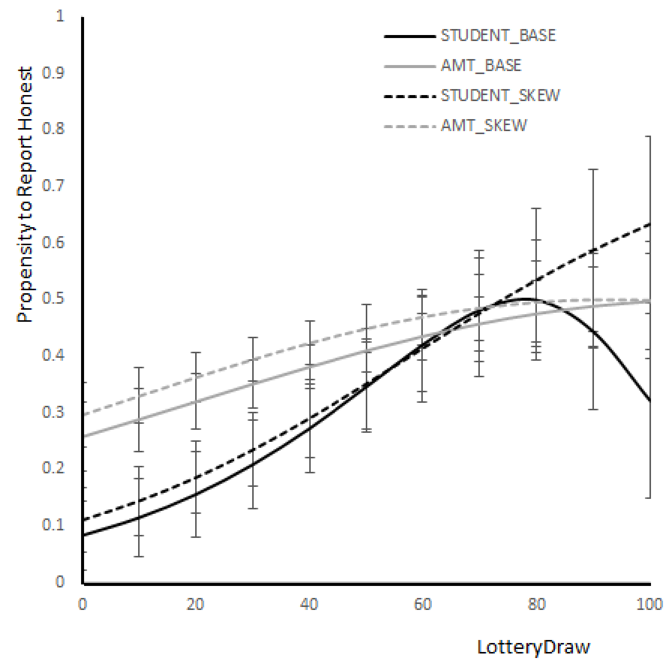


Figure 3.4: Propensity of subjects to report a value classified as Honest

or the SKEW treatment (FET: p = 0.343). Values for the test of the contrast of the propensity to report a DownLie value are given in Table C.4 in the Appendix. Notably no student subjects were observed to report a DownLie for low values of the LotteryDraw (under 78) in the BASE treatment whereas 13% of AMT subjects were observed to report a DownLie for the same range of values, leading to a significant difference for values of the LotteryDraw less than 70. There are no differences found in the propensity to report a DownLie between the samples in the SKEW treatment or between the treatments for either sample.

3.5.6 FullLie

The results presented in the previous sections are consistent with previous studies where subjects reported true values when presented with an incentive to report elevated false values that increased the payoff to both sender and receiver (Biziou-van-Pol et al. 2015; Erat and Gneezy 2012). However it is notable that a smaller proportion of student subjects made reports classified as Honest at low values of the LotteryDraw, and that a higher proportion reported values classified as an UpLie. Previous experimental results have presented evidence of partial lying, whereby subjects report false values that are not the payoff maximising ones (Abeler et al. 2016; Fischbacher and Föllmi-Heusi 2013). The observation of such reports is indicative of effects at the intensive margin, that is what value to report, as well as the extensive margin, that is the decision to report in a false manner. In this section reports classified as an UpLie are separated into two further categories of FullLie, whereby a subject is observed to falsely report the maximum value, and PositivePartialLie, where a subject falsely reports a value greater than the true value of the LotteryDraw, but that is not the maximum possible value of 100. A multinomial logistic model, again with a subject's choice of Honest chosen for the baseline, was performed using the alternative classifications. The results for the propensity to report an Honest value or a DownLie were not found to change significantly from the original model, so the discussion of these classifications is not repeated.

Result 3 *There are no significant differences in the tendency to report a FullLie between the samples. There is a lower propensity to report a FullLie by subjects in the AMT sample for values of the LotteryDraw of 40 or above in the SKEW treatment compared to the BASE treatment*

Figure 3.5 illustrates the proportions of different samples reporting a FullLie or a PositivePartialLie for the BASE and SKEW treatments. There are no significant differences in the proportion of subjects reporting a FullLie between the subject samples in either the BASE treatment (FET: $p = 1.000$) or the SKEW treatment (FET: $p = 0.844$). The proportion of subjects reporting a FullLie is higher in the BASE treatment compared to the SKEW treatment, though this difference is not significant for the student subjects (FET: $p = 0.496$) or the AMT subjects (FET: $p = 0.107$).

Figure 3.6 illustrates the calculated propensity to report a FullLie at various values of the LotteryDraw for the BASE and SKEW treatments among the student and AMT subjects. Values for the test of the contrast of the propensity to report a FullLie are given in Table C.6 in the Appendix. The contrast shows a significant difference in the propensity to report a FullLie between the BASE and SKEW

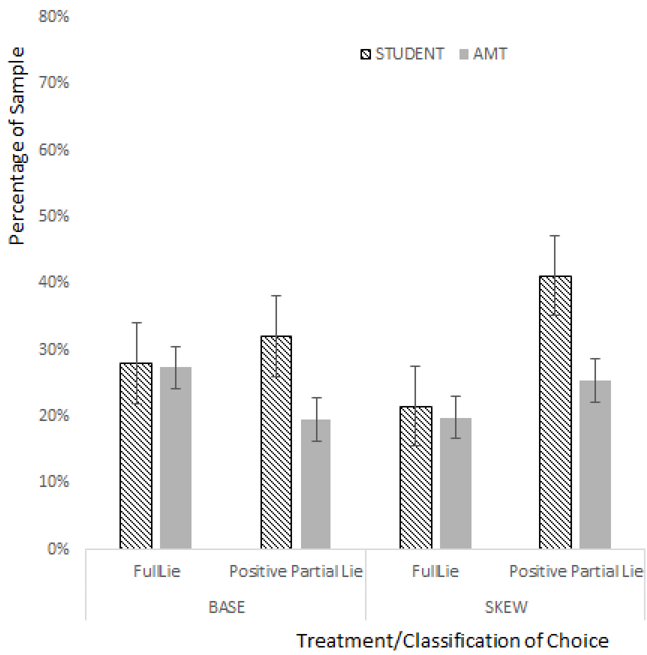


Figure 3.5: Proportion of subjects reporting a FullLie or PositivePartialLie

treatments for values of the LotteryDraw of 40 or above ($p \leq 0.0712$) among the AMT subjects, though there is no difference between the treatments found for the student subjects. This result is consistent with the second hypothesis.

3.5.7 PositivePartialLie

Result 4 *Student subjects demonstrated a higher tendency to report a PositivePartialLie than subjects recruited through AMT*

Figure 3.5 shows the proportion of the student sample reporting a value classified as a PositivePartialLie is higher in both the BASE (FET: $p = 0.082$) and SKEW (FET: $p = 0.036$) treatments than for the AMT subjects. This result can be linked with the observation of similar proportions in the two samples that that reported values classified as a FullLie in the previous result to note that the first result, a higher propensity of student subjects to report an UpLie, is driven by a higher tendency to report a PositivePartialLie among student subjects compared to AMT subjects.

Figure 3.7 illustrates the calculated propensity to report a PositivePartialLie at various values of the LotteryDraw. Values for the test of the contrast of the

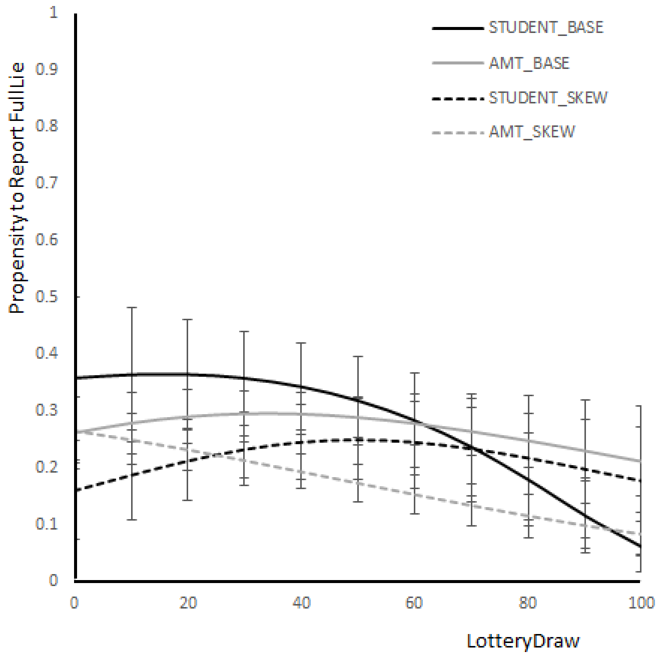


Figure 3.6: Propensity of subjects to report a value classified as a FullLie

propensity to report a PositivePartialLie value are given in Table C.7 in the Appendix. There is a significant difference in the calculated propensity to report a PositivePartialLie between the student and AMT subjects for values of the LotteryDraw of 30 or below ($p \leq 0.0284$) in the SKEW treatment and for values of the LotteryDraw in the range 30 to 60 inclusive $p \leq 0.0819$) in the BASE treatment. No differences were found in the propensity to report a PositivePartialLie between the treatments for either sample.

3.5.8 Controls

The evidence presented in the previous sections indicates that the student subjects reported more lies, in particular that there was a greater proportion found to report values that were classified as a PositivePartialLie. Tables C.1 and C.5 detail results from extended regressions on the classification of the reported value with additional control terms to check for correlations with personal characteristics declared by subjects and the subjects' actions within the experiment.

There is a clear negative correlation between the tendency to report an UpLie and Error, that is subjects who answered all the questions correctly on the quiz

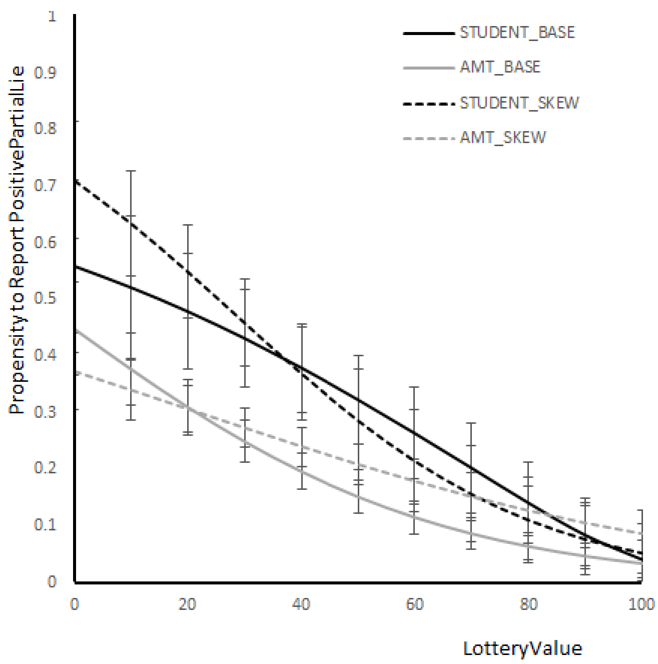


Figure 3.7: Propensity of subjects to report a value classified as a PositivePartialLie

were more likely to report a value larger than the true state. Conversely there is a positive correlation between the tendency to report a DownLie and Error, indicating that subjects who answered one or more questions incorrectly were more likely to report a value below the true value. The same pattern of a significant negative correlation with Error is found for the tendency to report a FullLie, and is negative but not significant for the case of a PositivePartialLie. Errors made by subjects on the quiz of understanding of the task are further discussed in the next subsection.

The length of time a subject spent reading the instructions (Lotteryinstructions) of the lottery task can be seen to positively correlated with the tendency to report a FullLie and negatively with to the tendency to report a DownLie. This suggests that the time spent reading the instructions may serve as a measure of the degree of attention that a subject gave to the experiment and in turn to the nature of the choice made. There is a negative correlation between the time spent answering the questions in the quiz of understanding of the experimental task (Lotteryquiz) and the tendency to report both an UpLie and a FullLie, further suggesting that the degree of understanding a subject has of the experiment influences the nature of their decision. The results also show a positive correlation between the time taken in the length of time to make a deicision of the value to report in the lottery task (Lotterydecision) and the tendency to report a PositivePartialLie, suggesting that the act of reporting a value that is a partial lie takes longer compared to reporting values of the other classifications. The role of the length of time subjects spent on the experiment is further discussed in a later section.

There are no effects found in relation to age or gender. There is a negative correlation observed between the tendency to report a DownLie and the measure of Extraversion, suggesting that people who agree more with the statement that they are reserved in nature are less likely to report a DownLie.

3.5.9 Errors

79.7% of the recruited sample (72.5% of the used sample) made at least one error on the quiz, such that it is clear that the subjects' level of understanding of the experimental task after reading the instructions is incomplete. The regression results and the analysis of timings presented in the preceding section detailed that there is positive correlation between making an error on the quiz and reporting a value classified as a DownLie as well as a negative correlation between time taken reading the instructions and the likelihood of reporting a DownLie. Both of these findings suggest that the choice to report a value below the true value may relate to a subject's lack of understanding of the experimental task. This section further investigates this link by examining the errors subjects made in the quiz of understanding of the experimental task.

There are no differences in the overall error rates between the treatments for either sample. Overall student subjects made fewer errors than AMT subjects (FET: $p = 0.022$), but the difference is weakly significant in the base treatment (FET: $p = 0.082$) and not significant for the SKEW treatment (FET: $p = 0.135$). Table 3.3 shows the results of logistic regressions on a binary measure as to whether or not a subject made an error in the quiz testing the subject's understanding of the lottery task. The estimates for two regressions are shown, the first for the set of subjects used in the main analysis, the second for all subjects who completed the final task in the experiment, therefore including those subjects who made more than one error on the quiz of understanding of the lottery task. The results show a negative correlation between the time spent reading the instructions and the propensity to make an error in the quiz, which is strengthened in the case where subjects excluded from the main analysis are included in the regression. The results also show that males were less likely to make an error and that there is no observable effect relating to the sample or the treatment. This result suggests that while student subjects proportionally recorded fewer errors on the quiz of understanding, this was in a large part due to the longer length of time they spent reading the instructions compared to AMT subjects.

DV: Error	Used Sample	All Completed
STUDENT	-0.5749	-0.3264
	(0.4751)	(0.432)
SKEW	-0.1199	-0.0443
	(0.2774)	(0.2719)
STUDENT SKEW	0.1647	0.0789
	(0.5174)	(0.4989)
Age	0.0003	0.0039
	(0.0155)	(0.0152)
Male	-0.5992**	-0.6398**
	(0.2678)	(0.267)
Extroversion	-0.1261*	-0.1204*
	(0.0672)	(0.0651)
Agreeableness	0.0462	0.0441
	(0.0685)	(0.065)
Emotional Stability	0.0417	0.0425
	(0.0629)	(0.061)
Conscientiousness	0.1103	0.1245*
	(0.0734)	(0.0717)
Openness	-0.0644	-0.0662
	(0.0702)	(0.0688)
Allocation	-0.012	-0.016
	(0.0509)	(0.0506)
Slider	-0.0219	-0.0232
	(0.016)	(0.0159)
Prelottery	-0.0003	-0.0002
	(0.0005)	(0.0004)
Lotteryinstructions	-0.0026*	-0.0025**
	(0.0015)	(0.0011)
Lotteryquiz	0.0036	0.0047
	(0.004)	(0.0044)

Lotterydecision	0.0047	0.0021
	(0.0061)	(0.0058)
cons	2.8289**	2.7161***
	(1.0524)	(0.9882)
N	461	517
Psudeo R2	0.07	0.07
LL	-229.0	-241.8

Table 3.3: Logistic estimates of determinants of making an error in the test of understanding

Figure 3.8 shows the proportion of subjects reporting by the given classifications separated by whether they made an error on the quiz as well by the treatment and sample. Consistent with the regression results presented in the previous section, the proportion of subjects reporting a DownLie is higher for subjects who made an error on the quiz compared to those who did not, and the proportion of subjects reporting a FullLie is lower. Overall a considerably lower proportion (3.8%) of subjects who made no error on the quiz of understanding reported a value classified as a DownLie compared to subjects (15.7%) who did make an error (FET: p = 0.001). The proportion reporting a value classified as a FullLie is higher among subjects who did not make a mistake on the quiz (40.9%) compared to those that did (16.4%) (FET: $p < 0.001$). The proportions reporting a value classified as an Honest response or as a PositivePartialLie do not differ between subjects who answered all the quiz questions correctly and those that did not.

Table 3.4 details the percentage of the used subjects who made an error on their first attempt at each of the questions in the test of understanding of the lottery task. Details of the questions are given in the Appendix. All of the questions in the quiz were multiple choice, with the default option set (incorrectly) as a blank. Questions 1, 2 and 4 were most frequently answered correctly by subjects, with 60.0% of subjects answering all three of these questions correctly (61.5% among student subjects and 59.5% among AMT subjects).

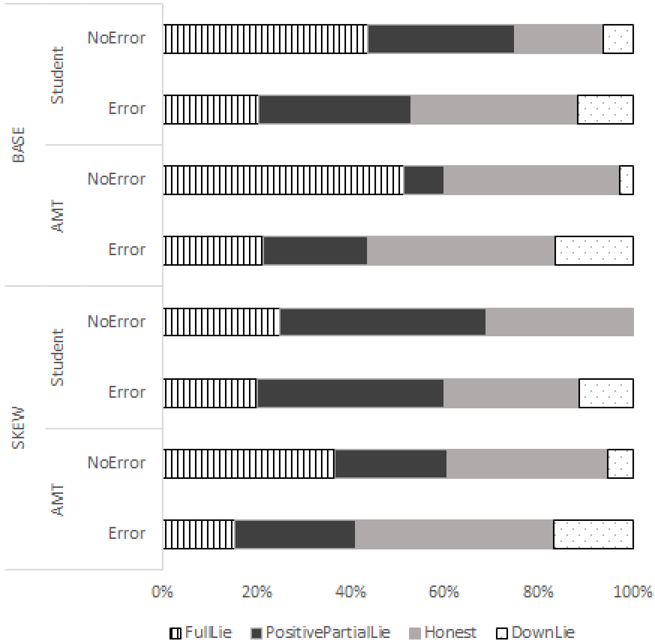


Figure 3.8: Proportion of subjects reporting by classification separated by Error

Sample	Trmt	Any	Q1	Q2	Q3	Q4	Q5	Q6
Student	BASE	68.0%	10.0%	26.0%	34.0%	6.0%	30.0%	26.0%
Student	SKEW	66.7%	5.9%	21.6%	43.1%	3.9%	29.4%	31.4%
AMT	BASE	78.8%	21.2%	11.7%	54.2%	10.1%	46.4%	50.3%
AMT	SKEW	69.1%	12.2%	18.2%	44.2%	10.5%	44.8%	40.9%

Table 3.4: Percentage of errors on first attempt at quiz of understanding of the lottery task of used sample

Question 4 tested if subjects knew the relative payoff to their partner player if they were chosen as player A, the correct response being 5. A large proportion of subjects answered this question correctly (89.6%) with no significant variation between samples or treatments, suggesting the majority of subjects had determined the key feature of the experiment. 49.9% of subjects, however, answered question 3 incorrectly, which checked if subjects understood the effect on payment of their choice if they were selected to be player B ("My value will not matter"). The proportion of student subjects answering this question incorrectly is lower than for the AMT subjects, particularly in the BASE treatment. A large proportion of the incorrect answers (91.9%) were 5, consistent with a view that subjects understood the important features of the experiment with regard to the nature of the payment, but potentially that a proportion were confused as to the nature of the roles detailed in the instructions.

A larger proportion of AMT subjects answered question 5, asking what the maximum payment available (£5 or $5), and question 6, asking for the minimum the AMT sample, paid attention to the example table and gave answers based on the table values rather than by calculating the appropriate values for questions 5 and 6.

The number of errors show that some subjects failed to fully understand the instructions for the lottery task. The nature of the errors, however, also reveals that the majority of subjects had particular misunderstandings that they were then able to correct the answers when prompted with the answers. While subjects who made a second mistake on the quiz have been excluded from the analysis, the evidence suggests that there remain some subjects who did not have a full understanding of the experiment and that the behaviour of such subjects is positively correlated to the tendency to report a DownLie and negatively with a FullLie but is not correlated with other the other classifications.

3.5.10 Experimental Timings

The correlation of times taken in certain steps within the experiment and the classification of outcome observed in the Controls section is worth further investigation. Spiliopoulos and Ortmann (2017) provide a summary of the growth of articles in experimental economics utilising response times in the examination of subjects decisions, as well making the case for using response time data in analysis. The authors suggest that the analysis of response times can be used in assisting with the classification of decision-makers into types and in investigating how decisions are made, in particular that the time observed by a subject in responding to a particular decision can be related to the nature of the choice. Piovesan and Wengström (2009) present evidence that faster subjects make more egotistic choices than slower ones. Rubinstein (2013) presents evidence demonstrating a link between short response times and decisions in a variety of games which are made in error. Chapter 1 of this thesis presents evidence of subjects taking a longer period of time to make responses that are non-compliant in a tax reporting experiment compared to those submitting compliant decisions, potentially arising from the additional cognitive complexity of deciding upon a value to misreport. Wang et al. (2010) present alternative experimental evidence using eye tracking and pupil dilation in a sender-receiver game that is consistent with a hypothesis that the process of deciding how much to deceive an opponent is cognitively complex. The analysis of the extended multinomial logistic regressions presented in the previous section indicated some interesting patterns of correlation between the subjects' choice and the time taken over various tasks within the experiment that are therefore further investigated in this section.

There are no significant differences in the distribution of times taken for any component of the experiment between the BASE and SKEW treatments within either of the student or AMT subjects (see Tables C.8 and C.9 in the Appendix). The recorded values of the times taken for each step over the experiment are therefore pooled across the treatments so as to be by sample in the following analysis.

Result 5 *Student subjects took significantly longer reading the instructions for all three parts of the experiment than subjects recruited through AMT*

Figure 3.9 shows the median time taken for the various steps within the experiment for the two sample pools. The median time taken for each of the steps is different between the two samples at a significance below the 1% level (Mann-Whitney (henceforth MW) tests). A longer median time was observed for student subjects (31 seconds) than for AMT subjects (22 seconds) to read the initial instructions for part one of the experiment (MW test: $z = 3.475$, $p < 0.001$). Likewise, a longer median time was observed for student subjects to read the instructions for the slider task (51 seconds) than for AMT subjects (39 seconds) (MW test: $z = 4.052$, $p < 0.001$). A longer median time was also observed for students to read the instructions for the final lottery based task of 244 seconds compared to 139.5 seconds for the AMT subjects (MW test: $z = 7.897$, $p < 0.001$). Subjects recruited through AMT were also observed to have shorter median times for completion of all the tasks in part 1 of the experiment.

AMT subjects, however, were observed to have a longer median value than student subjects for the time taken to complete the slider task and the quiz of

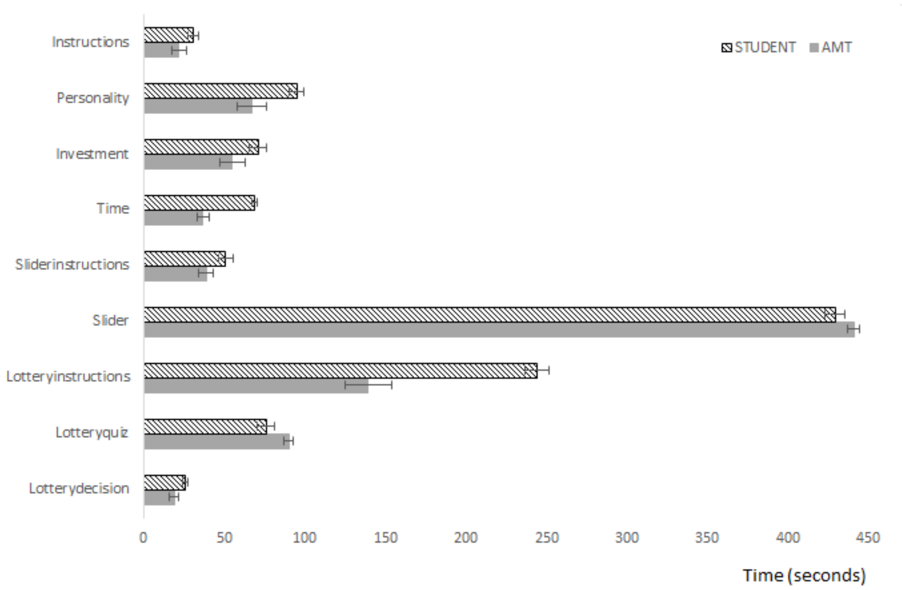
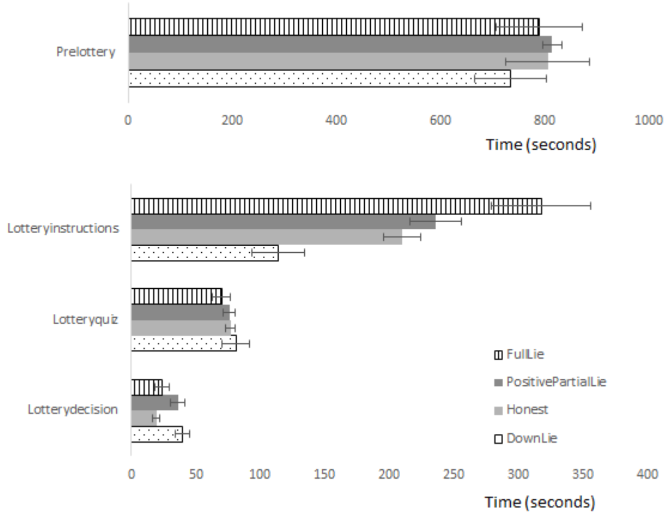


Figure 3.9: Median time taken for tasks within the experiment

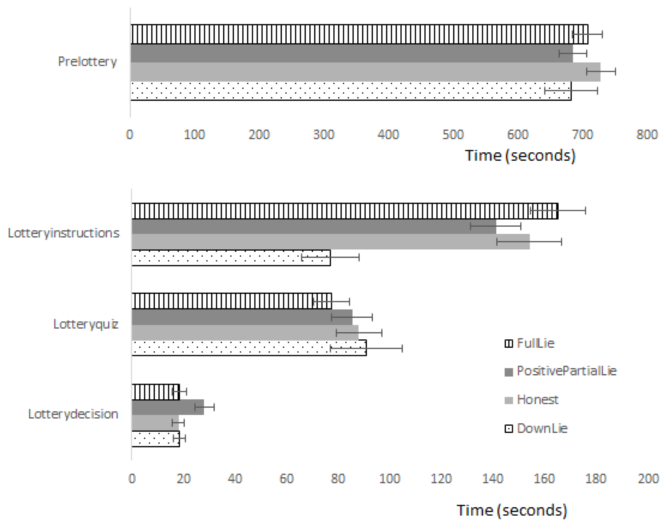
understanding of the lottery task. The longer median time observed for AMT subjects for the slider task may reflect a higher degree of network latency between the web server and the more geographically remote AMT subjects for this, technologically more burdensome, part of the experiment. The longer median time taken on the quiz checking for understanding of the lottery task may arise due to a higher rate of errors made by the AMT subjects, as discussed in the previous section.

Result 6 *Subjects who reported a DownLie spent a shorter time reading the instructions for the lottery*

Figure 3.10 shows the median length of time subjects spent on various steps of the experiment by sample and the classification of their reported value. There are no differences within the samples in the time spent on the experiment by classification prior to the lottery task (Prelottery). The median time spent reading the instructions for the lottery task (Lotteryinstructions) is far shorter among subjects who reported a DownLie than for subjects reporting any one of the three other categories among both student subjects (MW test: $z = 3.707$, $p < 0.001$) and AMT subjects (MW test: $z = 6.499$, $p < 0.001$). Figure 3.10 further shows there are no significant differences in the median length of time taken over the quiz



(a) Student Sample



(b) AMT sample

Figure 3.10: Median time taken for completion of steps of the experiment by classification

testing the subject's understanding of the experiment by classification in either sample.

Result 7 *AMT subjects take significantly longer in choosing a value to report when reporting a PositivePartialLie compared to a FulLie*

Figure 3.10 highlights that a longer median value was observed for the time taken to make a decision in the lottery task by AMT subjects who reported values classified as PostivePartialLie (Lotterydecisiontime) compared to reporting a FullLie (MW test: $z = 3.514$, $p < 0.001$). The median value observed for the time taken to report a PostivePartialLie by AMT subjects is also longer than that taken to report an Honest value (MW test: z = 2.704, p = 0.0068) and a DownLie (MW test: z = 2.326, p = 0.0200).

Figure 3.9 shows that student subjects were observed to have a longer median time (26 seconds) to choose a value to report than AMT subjects (19 seconds)(MW test: z = 2.615, p = 0.0089). Figure 3.10 shows that the median value for the time taken to make a decision in the lottery task was observed to be longer for student subjects reporting a value classified as a PositivePartialLie than for values classified as a FullLie or Honest, though the differences are not significant.

3.6 Discussion

The results presented in the previous section show that a higher proportion of student subjects were observed to report an UpLie, that is to misreport a lottery value for financial gain, than of subjects recruited through AMT. This result is consistent with the fist hypothesis that conjectured that AMT subjects would be less likely to misreport. There are, however, interesting features in the pattern of reporting between the two samples.

Similar proportions of subjects in the two samples were observed to report a FullLie in both the BASE or SKEW treatments, whereas a higher proportion of student subjects were found to report a PositivePartialLie compared to AMT subjects, particularly at lower values of the LotteryDraw in the SKEW treatment. A higher proportion of AMT subjects were observed to report an Honest value compared to student subjects over a corresponding range of low values for the LotteryDraw. These results suggest that the student subjects were more willing to lie than the AMT subjects given a sufficient incentive to lie. Interestingly, the primary difference is that such student subjects when willing to lie given a low value of the LotteryDraw, showed a greater tendency to report partial lies rather than the maximum value.

The results show that there are some significant differences in the personality measures between the samples, but that there is no particular correlation of these measures, or of gender, age or a measure of risk aversion to the tendency to report in a particular manner. The greater propensity for honesty and the associated lower tendency to report a positive lie may therefore reflect a greater concern for their reputation among the AMT subjects as hired workers rather than as pre-registered volunteers for research experiments as is true of the students. However recent research has shown that an extended version of the set of personality traits containing the additional trait of honesty-humility correlates better with dishonest behaviour (Hilbig and Zettler 2015) that was not collected in this study. The finding that AMT subjects exhibit a lower propensity to report false values and a higher propensity to report the truth is in keeping with the results of Choo et al. (2016), who found that a representative sample of taxpayers were more compliant than a sample of students, who claim their result is due to taxpayer's retaining norms of compliance in the experiment. The results is also similar to that of Abeler et al. (2014), who found that students where more likely to misreport the outcome of a series of coin tosses than members of the German public. A key difference of the study here is that AMT subjects were aware they were part of an experiment, as were the students, unlike the subjects from the German public.

There were no significant differences found in the proportions reporting for each of the classifications between the BASE and SKEW treatments for the student sample. A lower propensity to report a FullLie was observed in the SKEW treatment compared to the BASE treatment at higher values of the LotteryDraw for AMT subjects. These results are consistent with the second hypothesis that subjects should misreport no less frequently in the BASE treatment compared to the SKEW treatment. The result for the AMT subjects is stronger in that it provides some evidence in favour of the reference point model postulated in Chapter 2, in that more maximal lying is observed in the BASE treatment which has a higher reference point.

A relatively large proportion of subjects (13.0%) were observed to report a

DownLie despite the fact both subjects received a higher payment if the value reported was over stated from the truth. Though downward lies have been observed in one study (Utikal and Fischbacher 2013), a number of recent experiments have found little or no downward lying (Abeler et al. 2016; Gneezy et al. 2016), suggesting the relatively frequent observation of values classified as a DownLie in this study are a cause for concern.

The results show there is a positive correlation between the tendency to report a DownLie and subjects who made errors on the quiz of understanding of the lottery task. The results also show a negative correlation between subjects who made errors and the propensity to report a FullLie, but no correlation with the tendency to report the other categories. These observations suggest that there is some proportion of subjects who are able, when prompted with the solutions, to complete the quiz and progress to the decision step without a full understanding of the task. Such subjects are then more likely to report a DownLie and less likely to report a FullLie. Analysis of the time spent reading the instructions for the experiment shows that subjects reporting values as classified as a DownLie were observed to have spent a considerably shorter median time studying the instructions. These results suggest that some subjects did not fully understand the experiment, in a large part because they did not pay sufficient attention to the experiment and that such a lack of understanding lies behind a significant proportion of the choices classified as a DownLie.

While the literature review detailed number of studies that have replicated experiments using AMT subjects, a number of others have identified issues with the use of AMT subjects. Crump et al. (2013) suggest that testing subjects' comprehension of the task is critical, noting their results resembled the classical studies that they were attempting to replicate with subjects recruited through AMT when the data was analysed with the inclusion of such checks. Deetlefs et al. (2015) report an "unreliable" response level of 13% among subjects recruited through AMT, as well as a tendency for such "unreliable" subjects to rush through experiments. It is apparent from the analysis of experimental timings and the errors presented in this chapter that some proportion of the AMT sample pool do appear to conform to such “unreliable” behaviour. Notably, however, there is some proportion of the student subjects for which the same “unreliable” behaviour applies. An open question as to whether the observation of unreliable subjects arises from particular elements of the design used in this chapter or a wider problem for on-line experiments is left for future research.

Researchers are looking to alternative pools of subjects for experiments, to mitigate concerns relating to non-representative nature of student pools and practical issues such as the limited availability of subjects of laboratory space. Amazon Mechanical Turk serves as one of the recent wave of new sources of subjects for experimenters that addresses the specific issues of recruitment and payment. While a number of studies have replicated different aspects of student behaviour observed in the lab with AMT subjects on-line, this is the first study that I am aware of that compares the behaviour of AMT subjects with student subjects in deception game conducted on-line for both samples. While the main finding that a lower proportion of AMT subjects report self-serving false values is clear, there are also limitations in the analysis brought about by a fraction of subjects behaving in an “unreliable” manner. As increasing numbers of experimenters turn to use platforms such as AMT to source subjects, the results presented in this chapter give a clear indication that the potential for the presence of such responses must be considered.

Conclusion

The introduction to this thesis argued that people regularly tell little lies, but that people refrained from being dishonest at every opportunity, possibly due to some intrinsic motivation to adhere to a social norm of honesty, or potentially due to a preference to appear honest. The experimental results presented in each of the three chapters in this thesis are consistent with previous studies of dishonesty. A notable result is a propensity for honest responses even when a dishonest response is the rational choice, an outcome referred to as "lying aversion" in the literature (Gneezy 2005). Furthermore not all misreported values are largest lie possible, some are "partial lies" (Fischbacher and Föllmi-Heusi 2013), a well established experimental phenomena (Abeler et al. 2016). The three chapters of the thesis each examine a different aspect of the decision to be dishonest.

Chapter 1 detailed an experiment investigating the use of defaults in a tax filing specific context. The results showed that pre-population with incorrect values that favoured the taxpayer lead to an increased level dishonesty. While some proportion of the non-compliance was found to be passive, in the sense of only arising from the incorrectly pre-populated default value, the majority was found to be active, involving alterations to the value or further non-compliance in other fields. The chapter notes a number of future directions for research in the tax domain. An alternative avenue for future investigation could be to investigate the use a similar form of defaults in a neutral (non tax filing) frame. A potential example would be the use of a website to claim prizes from a lottery with a set of defaults corresponding to the prizes on offer. The results in the chpater also showed that the use of a prompt containing a descriptive norm was only effective in restoring compliance when used in a reactive manner. The effects of various static and dynamic nudges utilising various descriptive or injunctive norms, as discussed in the discussion of the chapter, could also be examined under such a framework.

Chapter 2 described how the results of experimental treatments with different reference points, in the sense of subjects being told about different modal values from previous sessions, lead to differences in dishonest behaviour consistent with a psychological model containing a cost in terms of the reference point. The chapter conjectured a number of different reasons for the existence of the additional cost, including a preference for the appearance of honesty, concern over the credibility of the value being reported and a desire not to boast. Chapter 3 provided a further result consistent with the model. The results from a sender-receiver game revealed that subjects had a higher propensity to report a maximal lie at higher values of the underlying draw in a treatment with a uniform distribution compared to one with a distribution skewed towards low values. There are two key avenues for future research.

The first avenue is to better establish the role for the distribution of underlying outcomes in the decision to be dishonest, much as in the manner of the recent experiments of Abeler et al. (2016) and Gneezy et al. (2016). One interesting alternative would to compare the responses subjects asked to undertake a draw with a uniform distribution, such as a six-sided dice roll, to those asked to undertake an equivalent draw with a binomial distribution, such as reporting the number of tails in 5 flips of a coin.

The second avenue to investigate further is the potential causes of the cost associated with reporting values above the reference point. The experiment described in this thesis establishes that there is a role for the reference point in relation to a norm of lying, the treatment whereby subjects are informed of higher reference point correlated to a greater propensity to misreport. The experiment, however, does not examine the roles of the potential causes of the change in behaviour, such as the credibility of the value declared or a preference for appearing hones. Abeler et al. (2016) detail an experiment designed to investigate the role of a reputation for appearing honest.

Chapter 3 adds to the literature on comparisons of behaviour between subject pools. The results show different patterns of dishonesty among students and workers from Amazon Mechanical Turk undertaking the same online experiment, a result consistent with those presented by Choo et al. (2016) and Abeler et al. (2014). An important finding of chapter 3 is the presence of a proportion of "unrealiable" subjects, with a tendency to rush through the experiment and not have a complete understanding of the experimental task in both pools of subjects. Such a result should serve as a warning to all experimenters using online experiments, whatever their source of subjects.

A result observed in both chapters 1 and 3 is that more complex deceitful messages, subject entered non-compliant levels of income in the experiment of chapter 1 and positive partial lies in the experiment of chapter 3, were correlated with longer median response times. This result matches to findings in the psychology literature, whereby more complex lies were found to be cognitively more demanding (Vrij and Heaven 1999) and to lead to longer response times (Vendemia et al. 2005). The results presented in this thesis back the conjecture of Spiliopoulos and Ortmann (2017) that response times can be useful in the understanding of decision makers' choices.

Appendix A

For Chapter 1

A.1 Participant Data

Treatment	Number Invited	Number Completed
BASE	105	80
CORR	109	88
OVER	109	83
UNDER	109	82
UNDERGENERIC	109	77
UNDERALWAYS	109	69
UNDERTRIGGER	109	75

[1]

Table A.1: Participant data

A.2 Experimental Materials

A.2.1 Screenshots

[1] Number Invited is the number of subjects invited to take part. Number Completed gives the number of subjects completing the treatment.

Instructions

Introduction

Welcome to our experiment. Please read these instructions carefully, as part of your payment will depend on the decisions you will make.

This is not a test, and there are no right or wrong ways to make your choices in this task.

In this experiment, you will take the role of a self-employed individual.

Your task in this experiment is to complete your annual tax return, and file it to an experimental tax authority. To do so, you will have to decide how much income and expenses to report, and thereby the amount of tax you choose to pay.

The task

You will be given a profile which breaks down all your sources of income and some of your expenses. One expense will not be given in the profile. You will generate this expense using a 6-sided dice. All items in your profile will be denominated in Experimental Currency Units (ECU). For each 1,000 ECU you accumulate, you will earn £0.50.

You will then need to complete a tax return , and declare to the experimental tax authority how much you have earned and how much you have spent .

You may find parts of your tax form have already been filled out by the experimental tax authority. The pre-filled values are estimates by the experimental tax authority of the values in your profile. These entries may not always be accurate. You can change the entries that are pre-filled if you wish to do so.

You will pay 40% in tax on the amount of income you declared minus whatever expenses you also declared. This will be calculated for you by the tax return.

After you finish your tax declaration, the experimental tax authority will decide whether to check your declaration.

The experimental tax authority will not know the contents of your profile. It can only find out the contents of your profile if your tax form is audited.

The decision of the experimental tax authority to check your tax return will depended on the values you enter. The chances of being audited depend on the tax liability (i.e. income-expenses) that you declare. The higher the tax liability you declare, the lower the probability that your return is checked will be. However, the chances that your return will be checked will never be higher than 10% (i.e. a 1-in-10 chance that the chosen profile is audited).

If you are audited, the experiment tax authority will access the information in your profile. The tax authority will know the actual values of income and expenses in your profile, except for any generated by your dice roll which cannot be audited. It will compare the amount you have declared with its own calculations.

If the audit determines you have under-declared your tax, you will pay the extra amount owed plus a fine. The fine will be 50% of the amount you under-declared by. That is, you have to pay an additional 5 ECU for every 10 ECU of unpaid tax.

If the tax authority made any mistakes when it pre-filled the tax return and you did not correct these, then these mistakes will also be discovered and the normal penalty applied.

In the following screen we will go through an example.

Figure A.1: Instructions Page 1

Instructions - 2

Example

Let's assume your actual income and expenses for the year are as follows:

Income	40,000
Expenses	2,000

The table below shows a few hypothetical examples of how you may choose to complete your tax return to the fictitious tax authority. For each example, the rows **Declared Income** and **Declared Expenses** reflect actual filing choices available to you.

- In example 1, you declared a different income to your earned income, but declared the same expenses as your incurred expenses.
- In example 2, you declared the same income to your earned income, but declared different expenses to your incurred expenses.
- In example 3, you declared the same income to your earned income, and declared the same expenses as your incurred expenses.

The bottom line in the table shows the final payment you're left with, depending on whether you're randomly selected for an audit or not.

	Example 1 (Under-declare income)		Example 2 (Over-declare expense)		Example 3 (Declare both accurately)
Income	40,000		40,000		40,000
Declared Income	**20,000**		**40,000**		**40,000**
Expenses	2,000		2,000		2,000
Declared Expenses	**2,000**		**4,000**		**2,000**
Taxable Income	18,000		36,000		38,000
Tax Due	7,200		14,400		15,200
Return Audited?	No	Yes	No	Yes	Yes or No
Additional Tax Payable	0	8,000	0	800	0
Penalty Payable	0	4,000	0	400	0
Payment (ECU)	**32,800**	**20,800**	**25,600**	**24,400**	**24,800**
Payment (£)	**16.40**	**10.40**	**12.80**	**12.20**	**12.40**

Your payment from taking part in this experiment therefore partly depends on how many ECUs you earn: this will be your income minus the tax you pay to the fictitious tax authority. The more ECUs you accumulate, the more you will earn. For every 1,000 ECU you accumulate, you will earn £0.50. At the end of the experiment your earnings will be converted into Pounds and paid to you.

Figure A.2: Instructions Page 2

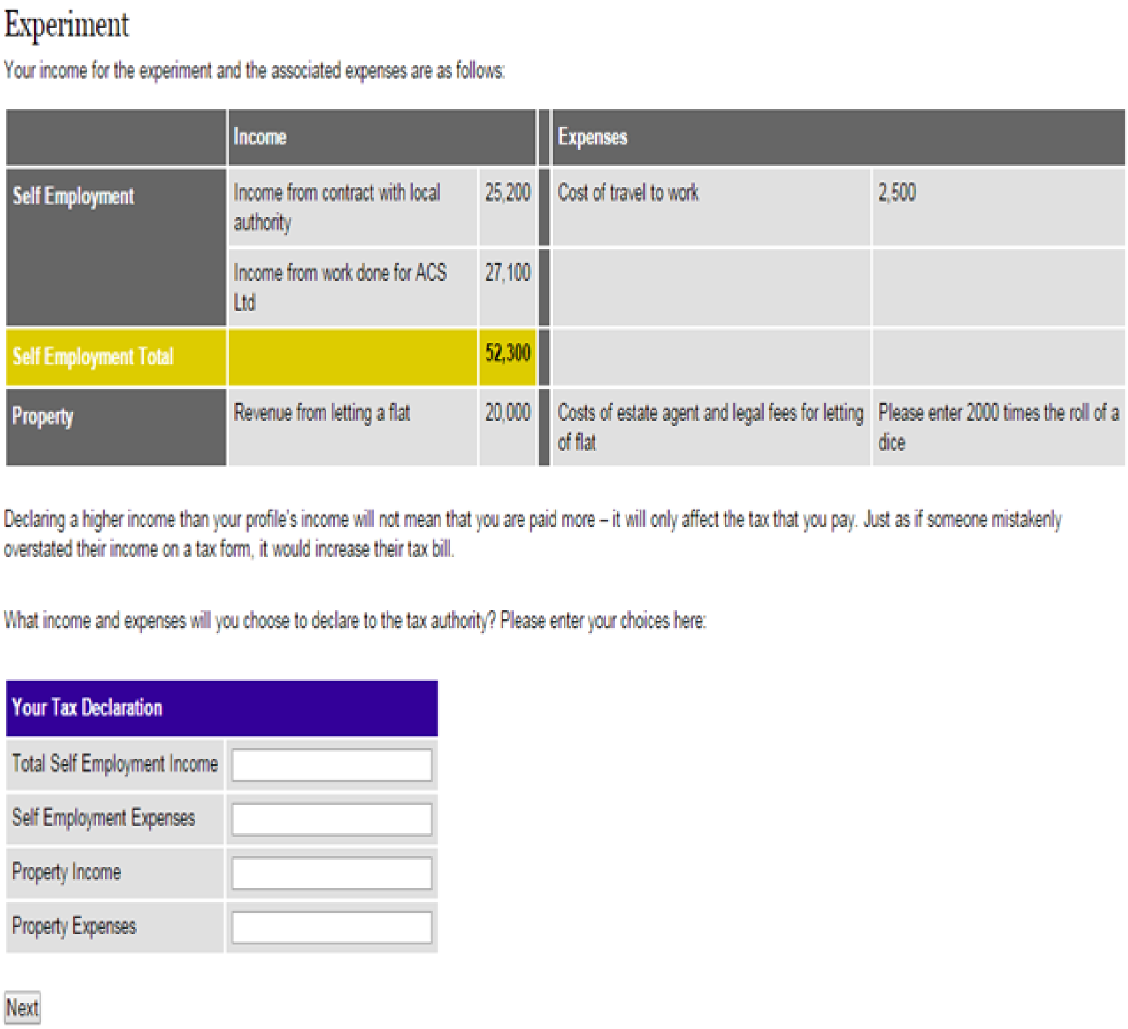
Instructions - 3

Summary

- You will be given a profile that includes different sources of income and expenses.
  - Your profile will include all of your sources of income.
  - Your profile will only include <u>some</u> of your expenses. One other will be generated by you using a 6-sided dice.
- You will be asked to complete a tax return.
  - You will pay 40% tax on your declared income after your declared expenses have been deducted.
  - The payoff will be based on the income shown in your profile minus any tax or penalty payments.
- You may find fields of your tax return might have already been filled out by the experimental tax authority:
  - Any pre-filled value is an estimate by the experimental tax authority of the field's true value in that profile
  - You can change the entries that are pre-filled if you wish to do so.
- The more income you declare, the more tax you will pay
- The more expenses you declare, the less tax you will pay
- Your payment for the experiment will be based on:
  - The profile's income and expenses
  - Your tax return for that profile and
  - Whether or not your return was selected for checking by the computer (with a maxiumum 1-in-10 chance).
- If your tax return is not checked, your payment will be your income minus tax
- If your tax return is checked, your payment will be:
  - Your income minus tax if the experimental tax authority finds no discrepancies; or
  - Your income minus tax plus a fine if the experimental tax authority finds discrepancies.

You will first carry out a practice round to help you further understand the format of the experiment. The values you enter in the practice round will not affect your end payment from the experiment.

Figure A.3: Instructions Page 3



Experiment

Your income for the experiment and the associated expenses are as follows:

	Income		Expenses	
Self Employment	Income from contract with local authority	25,200	Cost of travel to work	2,500
	Income from work done for ACS Ltd	27,100		
Self Employment Total		52,300		
Property	Revenue from letting a flat	20,000	Costs of estate agent and legal fees for letting of flat	Please enter 2000 times the roll of a dice

Declaring a higher income than your profile's income will not mean that you are paid more – it will only affect the tax that you pay. Just as if someone mistakenly overstated their income on a tax form, it would increase their tax bill.

What income and expenses will you choose to declare to the tax authority? Please enter your choices here:

Your Tax Declaration	
Total Self Employment Income	
Self Employment Expenses	
Property Income	
Property Expenses	

Next

Figure A.4: BASE treatment

Experiment

Your income for the experiment and the associated expenses are as follows:

	Income		Expenses	
Self Employment	Income from contract with local authority	25,200	Cost of travel to work	2,500
	Income from work done for ACS Ltd	27,100		
Self Employment Total		**52,300**		
Property	Revenue from letting a flat	20,000	Costs of estate agent and legal fees for letting of flat	Please enter 2000 times the roll of a dice

Declaring a higher income than your profile's income will not mean that you are paid more – it will only affect the tax that you pay. Just as if someone mistakenly overstated their income on a tax form, it would increase their tax bill.

What income and expenses will you choose to declare to the tax authority? Please enter your choices here:

Your Tax Declaration	
Information in tax authority database	25,200 and 27,100
Total Self Employment Income	52,300
Self Employment Expenses	
Property Income	
Property Expenses	

Next

Figure A.5: CORR treatment

Experiment

Your income for the experiment and the associated expenses are as follows:

	Income		Expenses	
Self Employment	Income from contract with local authority	25,200	Cost of travel to work	2,500
	Income from work done for ACS Ltd	27,100		
Self Employment Total		**52,300**		
Property	Revenue from letting a flat	20,000	Costs of estate agent and legal fees for letting of flat	Please enter 2000 times the roll of a dice

Declaring a higher income than your profile's income will not mean that you are paid more – it will only affect the tax that you pay. Just as if someone mistakenly overstated their income on a tax form, it would increase their tax bill.

What income and expenses will you choose to declare to the tax authority? Please enter your choices here:

Your Tax Declaration	
Information in tax authority database	25,200
Total Self Employment Income	25,200
Self Employment Expenses	
Property Income	
Property Expenses	

Next

Figure A.6: UNDER treatment

Experiment

Your income for the experiment and the associated expenses are as follows:

	Income		Expenses	
Self Employment	Income from contract with local authority	25,200	Cost of travel to work	2,500
	Income from work done for ACS Ltd	27,100		
Self Employment Total		52,300		
Property	Revenue from letting a flat	20,000	Costs of estate agent and legal fees for letting of flat	Please enter 2000 times the roll of a dice

Declaring a higher income than your profile's income will not mean that you are paid more – it will only affect the tax that you pay. Just as if someone mistakenly overstated their income on a tax form, it would increase their tax bill.

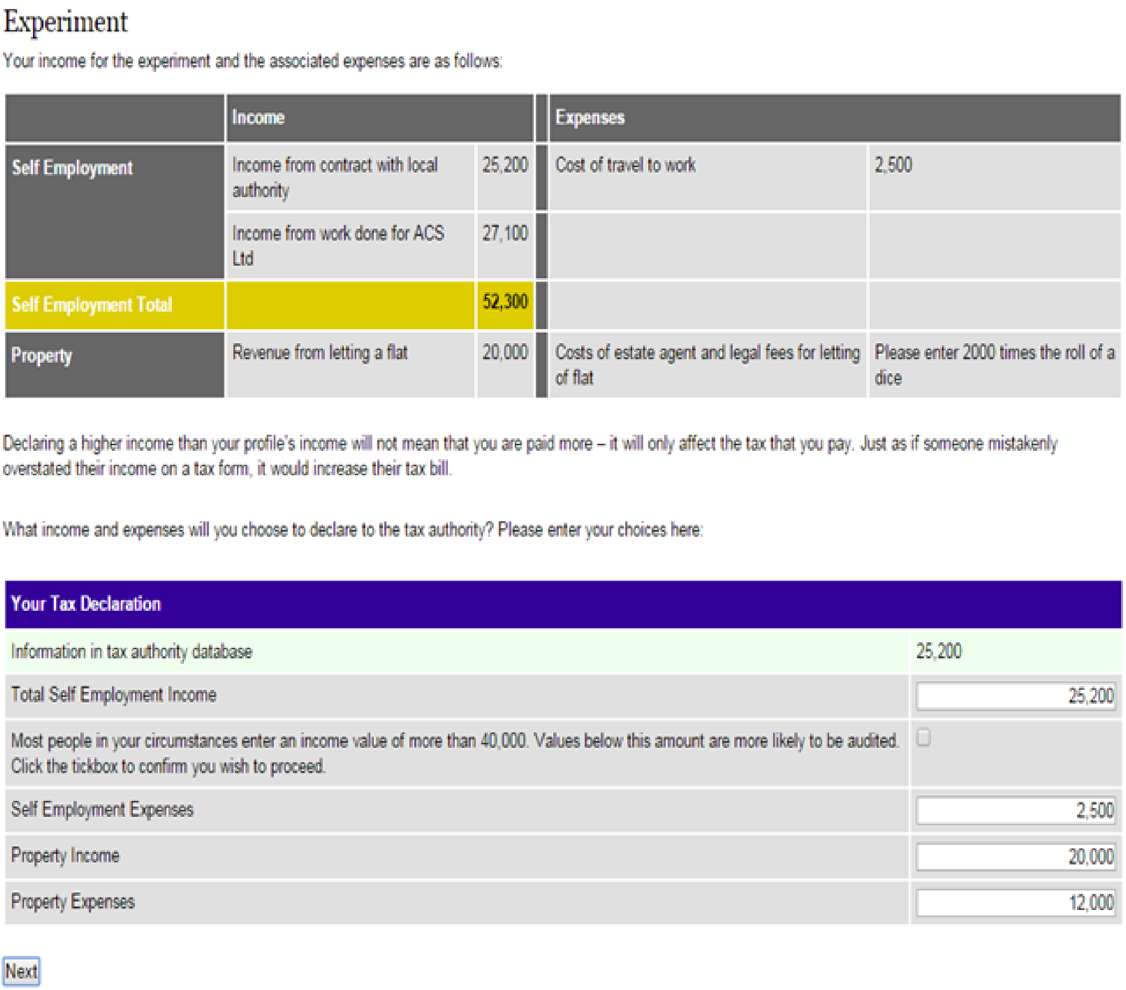
To change a pre-filled entry, please click on the check box next to the field.

What income and expenses will you choose to declare to the tax authority? Please enter your choices here:

Your Tax Declaration		
Information in tax authority database	25,200	
Total Self Employment Income	25,200	☑
Self Employment Expenses		
Property Income		
Property Expenses		

Next

Figure A.7: UNDERGENERIC treatment



Experiment

Your income for the experiment and the associated expenses are as follows:

	Income		Expenses	
Self Employment	Income from contract with local authority	25,200	Cost of travel to work	2,500
	Income from work done for ACS Ltd	27,100		
Self Employment Total		52,300		
Property	Revenue from letting a flat	20,000	Costs of estate agent and legal fees for letting of flat	Please enter 2000 times the roll of a dice

Declaring a higher income than your profile's income will not mean that you are paid more – it will only affect the tax that you pay. Just as if someone mistakenly overstated their income on a tax form, it would increase their tax bill.

What income and expenses will you choose to declare to the tax authority? Please enter your choices here:

Your Tax Declaration	
Information in tax authority database	25,200
Total Self Employment Income	25,200
Most people in your circumstances enter an income value of more than 40,000. Values below this amount are more likely to be audited. Click the tickbox to confirm you wish to proceed.	
Self Employment Expenses	2,500
Property Income	20,000
Property Expenses	12,000

Next

Figure A.8: UNDERTRIGGER treatment
2

Main Calculation

Based on the details you entered, your tax calculation is as follows:

Income	Value	Tax Payable
Employment Income	52,300	20,920
Property Income	20,000	8,000
Total Taxable Income	**72,300**	**28,920**
Expenses	**Value**	**Tax Relief**
Self Employment Cost of Goods	2,500	1,000
Self Employment Other Business Expenses	0	0
Property Expenses	12,000	4,800
Total Expenses	**14,500**	**5,800**
Tax Payable		**23,120**

You may now return and re-adjust your tax form or proceed to the end.

Re-enter Tax Next

Figure A.9: Tax calculation page

Appendix B

For Chapter 2

B.1 Proposition 1

Given that the set S is closed and compact and the utility function is continuous, then Weierstrass's theorem states that some non-empty compact set exists for the value to report to achieve the maximum utility for each value of the realised state. Denote this set as the correspondence $\mathrm{r} * (\mathrm{s}, \mathrm{s}^{\mathrm{R}})$ for each value of the state. For two realised states with different values, where $\mathrm{s}_2 > \mathrm{s}_1$, denote the minimum values of r that obtain the utility maximum as r_2 and r_1 respectively. That is $\mathrm{r}_2 = \min\{\mathrm{r}^*\mathrm{s}_2, \mathrm{s}^{\mathrm{R}}\}$ and $\mathrm{r}_1 = \min\{\mathrm{r}^*\mathrm{s}_1, \mathrm{s}^{\mathrm{R}}\}$. Given that the two values r_2 and r_1 yield maximal utility at their respective states, it follow that:

$$U(r_1, s_1, s^R) \geq U(r_2, s_1, s^R) \tag{B.1}$$

and

$$U(r_2, s_2, s^R) \geq U(r_1, s_2, s^R) \tag{B.2}$$

Using the specification of the utility function, these expressions can be rewritten as:

$$v(r_1) - c(r_1 - s_1) - c^R(r_1 - s^R) \geq v(r_2) - c(r_2 - s_1) - c^R(r_2 - s^R) \tag{B.3}$$

and

$$v(r_2) - c(r_2 - s_2) - c^R(r_2 - s^R) \geq v(r_1) - c(r_1 - s_2) - c^R(r_1 - s^R) \tag{B.4}$$

Summing over these two expressions yields:

$$-c(r_1 - s_1) - c(r_2 - s_2) \geq -c(r_2 - s_1) - c(r_1 - s_2) \tag{B.5}$$

which can be expressed as:

$$c(r_2 - s_1) - c(r_1 - s_1) \geq c(r_2 - s_2) - c(r_1 - s_2) \tag{B.6}$$

Given the definition that $s_2 > s_1$, then it follows $r_1 - s_1 > r_1 - s_2$ and $r_2 - s_1 > r_2 - s_2$. The proposition can be proved by a contradiction. Under a further assumption of $r_2 < r_1$, then the statements $r_1 - s_1 > r_2 - s_1$ and $r_1 - s_2 > r_2 - s_2$ must also hold, and therefore the conditions $r_2 - s_2 < r_1 - s_2 < r_1 - s_1$ and $r_2 - s_2 < r_2 - s_1 < r_1 - s_1$ must also be true. Ignoring temporarily the trivial case where $c(r - s) = 0$ everywhere, from these expressions we can use the convexity assumption on $c(r - s)$ to write:

$$c(r_1 - s_2) < \frac{s_2 - s_1}{r_1 - s_1 - r_2 + s_2} c(r_2 - s_2) + \frac{r_1 - r_2}{r_1 - s_1 - r_2 + s_2} c(r_1 - s_1) \tag{B.7}$$

and

$$c(r_2 - s_1) < \frac{r_1 - r_2}{r_1 - s_1 - r_2 + s_2} c(r_2 - s_2) + \frac{s_2 - s_1}{r_1 - s_1 - r_2 + s_2} c(r_1 - s_1) \tag{B.8}$$

Summing over these two expressions gives:

$$c(r_1 - s_2) + c(r_2 - s_1) < c(r_2 - s_2) + c(r_1 - s_1) \tag{B.9}$$

which can be expressed as:

$$c(r_2 - s_1) - c(r_1 - s_1) < c(r_2 - s_2) - c(r_1 - s_2) \tag{B.10}$$

However, equation B.10 is a contradiction to the statement in equation B.6. As the optimal value to report is always one in the case of no cost, we can therefore state that $r_2 \geq r_1$.

B.2 Proposition 3

For two reference points where $s_2^R > s_1^R$, denote the minimum values of r that obtain the utility maximum as r_2 and r_1 respectively. It follows that:

$$U(r_1, s, s_1^R) \geq U(r_2, s, s_1^R) \tag{B.11}$$

and

$$U(r_2, s, s_2^R) \geq U(r_1, s, s_2^R) \tag{B.12}$$

Using the specification of the utility function, these expressions can be rewritten as:

$$v(r_1) - c(r_1 - s) - c^R(r_1 - s_1^R) \geq v(r_2) - c(r_2 - s) - c^R(r_2 - s_1^R) \tag{B.13}$$

and

$$v(r_2) - c(r_2 - s) - c^R(r_2 - s_2^R) \geq v(r_1) - c(r_1 - s) - c^R(r_1 - s_2^R) \tag{B.14}$$

Summing over these two expressions yields:

$$-c^R(r_1 - s_1^R) - c^R(r_2 - s_2^R) \geq -c^R(r_2 - s_1^R) - c^R(r_1 - s_2^R) \tag{B.15}$$

which can be expressed as:

$$c^R(r_2 - s_1^R) - c^R(r_1 - s_1^R) \geq c^R(r_2 - s_2^R) - c(r_1 - s_2^R) \tag{B.16}$$

Given the definition that $s_2^R > s_1^R$, then it follows $r_1 - s_1^R > r_1 - s_2^R$ and $r_2 - s_1^R > r_2 - s_2^R$. To test for differences, if we further assume $r_2 < r_1$, then the statements

$r_1 - s_1^R > r_2 - s_1^R$ and $r_1 - s_2^R > r_2 - s_2^R$ must also hold. This leads to the observations that $r_2 - s_2^R < r_1 - s_2^R < r_1 - s_1^R$ and $r_2 - s_2^R < r_2 - s_1^R < r_1 - s_1^R$. Ignoring temporarily the trivial case where $c^R(r - s^R) = 0$ everywhere, from these expressions we can use the convexity assumption on $c^R(r - s^R)$ to write:

$$c^R(r_1 - s^R2) < \frac{s_2^R - s_1^R}{r_1 - s_1^R - r_2 + s_2^R} c^R(r_2 - s_2^R) + \frac{r_1 - r_2}{r_1 - s_1^R - r_2 + s_2^R} c^R(r_1 - s_1^R) \quad \text{(B.17)}$$

and

$$c^R(r_2 - s_1^R) < \frac{r_1 - r_2}{r_1 - s_1^R - r_2 + s_2^R} c(r_2 - s_2^R) + \frac{s_2 - s_1^R}{r_1 - s_1^R - r_2 + s_2^R} c^R(r_1 - s_1^R) \quad \text{(B.18)}$$

Summing over these two expressions gives:

$$c^R(r_1 - s_2^R) + c(r_2 - s_1^R) < c^R(r_2 - s_2^R) + c^R(r_1 - s_1^R) \quad \text{(B.19)}$$

which can be expressed as:

$$c^R(r_2 - s_1^R) - c(r_1 - s_1^R) < c(r_2 - s_2^R) - c(r_1 - s_2^R) \quad \text{(B.20)}$$

This is a contradiction to the statement in equation B.16. As the optimal value to report is always one in the case of no cost, we can therefore state that $r_2 \geq r_1$.

B.3 Change of Reference Point

The illustration in Figure 2.4 illustrates a slightly different case to that given in the main body of the text in that the set of states, S, has been extended to cover the range $[-1, 1]$. This does not affect the basic results. Setting the reference point at $s = 0$ under this extended range does however simplify the analysis under a change in the reference point, as alternative reference points can be viewed as ranges within the extended space, such that for any given refrence point, $s^R \in [0, 1]$, the appropriate sub range of the extended space would be $[-s^R, 1 - s^R]$.

An increase in the reference point from s^R to $s^{R'}$ can therefore be viewed as a left diagonal shift in the space under analysis. Such a shift is illustrated in the left panel of Figure 2.4. For any state in the actual range with the lower reference point, that is some $x - s^R \in [-s^R, 1 - s^R]$ where $x \in [0, 1]$, there is a corresponding value in the range with the higher reference point, $x' - s^{R'} \in [-s^{R'}, 1 - s^{R'}]$ such that $x = x'$. Under these conditions, the values of x and x' represent the same real state in the actual range under consideration for the two different reference points. The optimal reported value for the actual realised states for the two reference points is shown in the right hand panel of Figure 2.4, taken from the overlay of the two ranges in the left hand panel on the range of actual values $[0, 1]$.

B.4 Experimental materials

B.4.1 Desk Configuration

Figure B.1: Laboratory desk configuration prior to experimental session

B.4.2 Part One

The booklet used for part one of the experiment is reproduced on the next six pages, this includes the instructions read aloud at the start of the experiment.

B.4.3 Part Two

The booklet used for part one of the experiment is reproduced on the two pages that following after that, this includes the instructions read aloud at in-between the components of the experiment.

Welcome to today's experiment.

Please do not write your name or student id on this form to ensure anonymity.

This is an experiment on economic decision making, there are no right or wrong choices.

You will be paid £3 for showing up today and may have the opportunity to earn more money through your choices in the experiment. It is important that your actions in this experiment are anonymous to the experimental team. As the university finance department require that payments are recorded you will still be required to sign for your payment, but this record will only be passed to the finance department and the experimentalists will have no access to these records. **Your payments will therefore be anonymous to the experimental team.** Please bear in mind that the experiment is designed to be anonymous and do not use information that would indicate your identity, such as your name or student id, in any of the materials in the laboratory other than the PARTICIPANT PAYMENT RECEIPT form.

The experiment will be in two parts, we will introduce the second part later in the session. In the first part of the experiment you will be required to complete a questionnaire which is attached to these instructions. We will allow 10 minutes for you to complete the questions attached. If you have any questions, please raise your hand and the experimenter will come to you to address it.

Experiment Part 1 Questionnaire – Please Complete ALL Sections

Please do not write your name or student id on this form to ensure anonymity.

SECTION 1

Please complete the following questions where appropriate (or leave blank if you do not wish to answer)

Q1] Please let us know your gender: Male [ ] or Female [ ]

Q2] Please let us know your age: ______

Q3] Please let us know your course of study: ________________________

Q4] Please let us know what year of your studies are you in: _______________________

Q5] Please tell us about your household income, is it

> more than £10,000 [ ]

> more than £20,000 [ ]

> more than £40,000 [ ]

> more than £60,000 [ ]

> more than £100,000 [ ]

None of the above [ ]

Q6] Below are five pairs of descriptions. Circle **one point** on each scale to indicate how much you think each description sounds like you. For example:

- If a pair of descriptions describe you equally well, then mark the centre of the scale

Description 1 |-|-|-|-|-|-|-|-| **Description 2**

- If you are slightly more like description 1 than description 2, then mark the scale slightly closer to description 1

Description 1 |-|-|-|-|-|-|-|-| **Description 2**

- If description 2 is exactly right and description 1 is not like you at all, then mark the scale right next to description 2

Description 1 |-|-|-|-|-|-|-|-| **Description 2**

How much does each description sound like you? Generally I come across as:

Someone who is talkative, outgoing, is comfortable around people, but could be bossy and attention seeking	|-|-|-|-|-|-|-|-|	Someone who is a reserved, private person, doesn't like to draw attention to themselves and can be shy around strangers
Someone who is forthright, tends to be critical and find fault with others and doesn't suffer fools gladly	|-|-|-|-|-|-|-|-|	Someone who is generally trusting and forgiving, is interested in people, but can be taken for granted and finds it difficult to say no
Someone who is sensitive and excitable, and can be tense	|-|-|-|-|-|-|-|-|	Someone who is relaxed, unemotional rarely gets irritated and seldom feels blue
Someone who likes to plan things, likes to tidy up, pays attention to details, but can be rigid or inflexible	|-|-|-|-|-|-|-|-|	Someone who doesn't necessarily work to a schedule, tends to be flexible, but disorganised and often forgets to put things back in their proper place
Someone who is a practical person who is not interested in abstract ideas, prefers work that is routine and has few artistic interests	|-|-|-|-|-|-|-|-|	Someone who spends time reflecting on things, has an active imagination and likes to think up new ways of doing things, but may lack pragmatism

SECTION 2

Please consider the following hypothetical investment.

This investment has a 50/50 chance of success. If successful, any amount invested will return 2.5 times the value invested. However, if not successful then it will return nothing. For example, if all £10 were to be invested, then if it is successful, it will return £25, but if unsuccessful it will return £0.

You may invest any portion of the £10 (from 0 to 10), and keep the remainder.

Suppose you invested £3, and kept the remaining £7.
If the investment is successful, then you will receive £7.50 from the investment, plus £7 that you kept for a total of £14.50. However, if the investment is not successful, then you will receive nothing from the investment and only retain the amount you kept for a total of £7.

Alternatively, suppose you invested £8 and kept the remaining £2.
If the investment is successful, then you will receive £20 from the investment, plus £2 that you kept for a total of £22. However, if the investment is not successful, then you will receive nothing from the investment and only retain the amount you kept for a total of £2.

How much of the £10 would you invest?_____

SECTION 3

In this section you are asked to choose from a list of hypothetical options corresponding to different amounts of money to be paid in the future. Which is the first of the following options that you would select in preference to receiving £100 in one month's time:

Option 1: £100.00 in thirteen months' time

Option 2: £101.50 in thirteen months' time

Option 3: £103.00 in thirteen months' time

Option 4: £104.50 in thirteen months' time

Option 5: £106.00 in thirteen months' time

Option 6: None of the above

So if the first option you would prefer is £104.50 in thirteen months' time to £100 in one month's time, (such that you would not have preferred £100, £101.50 or £103.00 in thirteen months' time), you would select option 4.

Please select which option you would choose.

I would choose option: ___

SECTION 4

The purchase of a bat and a ball cost £1.10 (one pound and ten pence). The bat cost £1 more than the ball. How much did the ball cost?

The ball cost ____ pence

Experiment Part 2 Numeracy Test

In the second part of the experiment you have the opportunity to increase your payment for today's experiment in a numeracy test. Once instructed to turn over you will find 20 questions of the following form on the reverse:

3+5-2=

You will be given 30 seconds to try and complete as many of the 20 questions as possible.

You will earn 20 pence for each of the questions you complete.

You will be told when the 30 seconds is to begin and when it is at an end.

Once the 30 seconds is over you should check how many questions you have answered and work out your payment. You have been provided on the desk with a small white envelope containing 20 twenty pence pieces. You should remove the appropriate number of twenty pence pieces from the envelope for your payment. You should then seal any remaining coins in the small white envelope.

In previous sessions using this test one of the two most commonly reported values was 10 questions.

You should then write the value of you payment PLUS the £3 for you show up fee on your PARTICIPANT PAYMENT RECEIPT form. For example, if you complete 11 puzzles, you should remove 11 twenty pence pieces corresponding to £2.20 to add to your £3 show up fee. You should then complete the PARTICIPANT PAYMENT RECEIPT form for the total of £5.20.

You should then place all the experimental materials, including the booklets for the first and second part of the experiment as well as the small white envelope containing any remaining coins into the large plastic envelope provided, BUT NOT the PARTICIPANT PAYMENT RECEIPT which you should keep separate.

You will then be asked to leave. As you leave, there will be two boxes by the door. Please place the plastic envelope containing the experimental materials in the large left hand box and the place the PARTICIPANT PAYMENT RECEIPT face down in the box on the right.

PLEASE WAIT UNTIL YOU ARE ASKED TO START THE NUMERACY TEST BEFORE TURNING OVER THIS SHEET

1. 1+3-2=____

2. 4+5-7=____

3. 5+9-4=____

4. 3+7-2=____

5. 8+4-7=____

6. 9+4-6=____

7. 3+1-2=____

8. 5+6-3=____

9. 6+2-7=____

10. 8+9-3=____

11. 9+4-6=____

12. 5+3-7=____

13. 6+6-4=____

14. 4+3-2=____

15. 5+6-7=____

16. 7+3-4=____

17. 9+8-3=____

18. 3+8-5=____

19. 7+3-8=____

20. 6+9-7=____

B.5 Regression Results

DV: ReportedCalculations	Model 1	Model 2	Model 3	Model 4	Model 5
cons	2.8602	2.2268	3.6233	2.962	2.5765
	(2.7037)	(2.6745)	(3.1828)	(3.6231)	(3.5656)
RCAR	21.407***	21.448***	20.628***	24.9234***	24.5728***
	(3.2856)	(3.249)	(3.2556)	(4.8991)	(4.7356)
CCAR	-24.101***	-24.1697***	-23.8077***	-26.7994***	-26.6172***
	(2.8831)	(2.8479)	(2.8708)	(3.6587)	(3.5095)
LOW	2.6429	2.2861	2.3505	2.1723	1.8443
	(3.8461)	(3.792)	(3.7911)	(5.4215)	(5.3274)
LOW RCAR	2.1533	2.5936	2.2111	1.6453	2.2517
	(5.0332)	(5.0215)	(4.9992)	(7.2993)	(7.1972)
LOW CCAR	-4.4308	-3.4682	-3.1745	-3.9333	-3.3478
	(4.4307)	(4.4435)	(4.4737)	(5.3705)	(5.3168)
HIGH	-1.076	0.0019	-0.9185	-1.2637	-0.5046
	(5.1913)	(5.131)	(5.0584)	(6.7128)	(6.6103)
HIGH RCAR	-1.9699	-2.9534	-1.9583	-4.1211	-4.6388
	(5.743)	(5.6762)	(5.6198)	(7.8838)	(7.7364)
HIGH CCAR	7.4679**	7.255**	7.3093*	10.1335**	9.6931**
	(3.7615)	(3.7085)	(3.7381)	(4.6506)	(4.5333)
CorrectCalculations	0.7097*	0.7673*	0.6857*	0.6946	0.7242
	(0.3999)	(0.3946)	(0.3936)	(0.5096)	(0.5015)
RCAR CorrectCalculations	-2.1495**	-2.1892**	-2.0952**	-2.5928***	-2.5709***
	(0.4745)	(0.4692)	(0.47)	(0.661)	(0.6412)
CCAR CorrectCalculations	2.5633**	2.5604**	2.5097**	2.9325***	2.8966***
	(0.3055)	(0.3025)	(0.3065)	(0.434)	(0.415)
LOW CorrectCalculations	-0.3512	-0.2976	-0.3005	-0.2819	-0.2327
	(0.5659)	(0.558)	(0.5562)	(0.7492)	(0.7369)
LOW RCAR CorrectCalculations	-0.1626	-0.2358	-0.1963	-0.0807	-0.1755
	(0.714)	(0.7107)	(0.7065)	(0.9729)	(0.9596)
LOW CCAR CorrectCalculations	0.4349	0.3733	0.3436	0.3282	0.307
	(0.4894)	(0.4918)	(0.499)	(0.6389)	(0.6315)
HIGH CorrectCalculations	0.087	-0.0461	0.0303	0.1135	0.0218
	(0.7159)	(0.7072)	(0.6984)	(0.8865)	(0.8733)
HIGH RCAR CorrectCalculations	0.6797	0.8011	0.6848	0.9992	1.0602
	(0.8012)	(0.7915)	(0.7857)	(1.0449)	(1.0258)
HIGH CCAR CorrectCalculations	-1.0165**	-1.0106**	-0.987**	-1.3867**	-1.3525**
	(0.4186)	(0.4132)	(0.4164)	(0.5726)	(0.5572)
Male		1.0711***	0.8662*		0.8129**
		(0.4002)	(0.4495)		(0.3446)
Age			-0.0128		
			(0.0433)		
Allocation			0.062		
			(0.1006)		
Option			-0.109		
			(0.1925)		
CRTCorrect			0.1791		
			(0.476)		
Extroversion			-0.0303		
			(0.0969)		
Agreeableness			0.0248		
			(0.1062)		
Emotional Stability			0.0275		
			(0.1055)		
Conscientiousness			-0.0712		
			(0.0873)		
Openness			-0.0368		
			(0.1058)		
Economics			0.7348		
			(0.5054)		
AccountingFinance			1.7295		
			(1.0674)		
STEM			-0.2228		
			(0.6034)		
sigma	2.9151	2.8723	2.8102	1.2238	0.8451
alpha0				1.5855	1.9315
alpha1				-0.0746	-0.0734
N	258	258	258	258	258
Pseudo R2	0.21	0.22	0.22	0.23	0.23
LL	-539.9	-536.3	-530.4	-528.4	-525.6
Intercept+Slope:SW	0.7342	0.8651	0.6868	0.8634	0.9171
Intercept+Slope:NW	0.0097***	0.0086***	0.0227**	0.0205**	0.016**
Intercept+Slope:NE	0.0090***	0.0224**	0.0423**	0.0018	0.0046***

[1]

Table B.1: Censored regression estimates of determinants of number of ReportedCalculations

[1]Model 1 as specified by Equation 2.15, Model 2 incorporates the variable Male, Model 3 incorporates other subject responses, Model 4 uses Equation 2.15 and Equation 2.16 for heteroskedasticity, Model 5 incorporates the variable Male

		CorrectCalculations < 10		CorrectCalculations ≥ 10	
	Treatment	Intercept	Slope	Intercept	Slope
ReportedCalculations ≥ 10	LOW	29.0634	-1.9536	0.5317	1.0446
		(2.6751)	(0.3546)	(2.0901)	(0.1510)
	HIGH	21.2213	-0.6732	4.5883	0.8736
		(1.6123)	(0.2545)	(1.8368)	(0.1359)
ReportedCalculations < 10	LOW	5.5031	0.3585		
		(2.7354)	(0.4004)		
	HIGH	1.7842	0.7967		
		(4.4317)	(0.5938)		

Table B.2: Calculated coefficient values for censored regression

	FullFalse		PartialFalse	
DV: classif	Model L1	Model L2	Model L1	Model L2
CorrectCalculations	-0.3515***	-0.4294***	-0.2006***	-0.2263***
	(0.1092)	(0.1281)	(0.0635)	(0.0699)
LOW	2.5101	1.78470.323	1.1198	1.1098
	(1.6102)	(1.8064)	(1.1955)	(1.2858)
HIGH	-0.6023	-0.61670.652	-0.9355	-0.9958
	(1.2242)	(1.3688)	(1.0172)	(1.0639)
LOW CorrectCalculations	-0.299	-0.22170.265	-0.0856	-0.097
	(0.1879)	(0.1988)	(0.1001)	(0.109)
HIGH CorrectCalculations	0.1941	0.16940.221	0.0621	0.0601
	(0.1257)	(0.1384)	(0.0868)	(0.0913)
Male		1.2422**		0.4125
		(0.5169)		(0.375)
Age		0.1354		0.1308
		(0.1275)		(0.1199)
Allocation		0.078		0.0481
		(0.1145)		(0.0873)
Option		0.0823		0.0384
		(0.2329)		(0.1493)
CRTCorrect		0.4211		-0.0805
		(0.5709)		(0.3749)
Extroversion		-0.1734		-0.0753
		(0.1145)		(0.0809)
Agreeableness		0.1202		0.1493*
		(0.1209)		(0.0883)
Emotional Stability		-0.0278		0.0119
		(0.1242)		(0.0862)
Conscientiousness		0.0052		0.0461
		(0.1077)		(0.0737)
Openness		-0.0245		0.0076
		(0.1227)		(0.0888)
Economics		0.8656		-0.0238
		(0.5878)		(0.4242)
AccountingFinance		1.8956*		0.0853
		(1.0311)		(0.8245)
STEM		-0.4494		-0.3487
		(0.7856)		(0.49)
Cons	2.2489**	-1.389	2.5368***	-0.9686
	(0.9739)	(3.336)	(0.7156)	(2.777)
N	258	258		
Pseudo R2	0.15	0.21		
LL	-227.5	-209.1		

[2]

Table B.3: Multinomial logistic estimates of False reporting

[2]Model L1 based on Equation 2.19, Model L2 incorporates other subject responses

CorrectCalculations	χ^2	p
0	1.55	0.2125
1	0.96	0.3274
2	0.46	0.4969
3	0.1	0.7525
4	0.03	0.8711
5	0.63	0.4272
6	2.43	0.1188
7	5.64	0.0176
8	9.98	0.0016
9	14.99	0.0001
10	19.75	0
11	22.7	0
12	22.71	0
13	20.24	0
14	16.71	0
15	13.26	0.0003
16	10.36	0.0013
17	8.08	0.0045
18	6.34	0.0118
19	5.02	0.025
20	4.02	0.045

Table B.4: Post-estimation contrast of the propensity to report a FullFalse value between HIGH and LOW treatments

CorrectCalculations	χ^2	p
0	2.32	0.1276
1	2.53	0.1117
2	2.66	0.1028
3	2.66	0.103
4	2.48	0.1156
5	2.12	0.1458
6	1.63	0.2016
7	1.1	0.2937
8	0.61	0.4363
9	0.21	0.6478
10	0.01	0.9396
11	0.13	0.7204
12	0.63	0.4261
13	1.36	0.2438
14	2.04	0.1528
15	2.56	0.1093
16	2.92	0.0874
17	3.14	0.0763
18	3.25	0.0716
19	3.24	0.0719
20	3.13	0.0768

Table B.5: Post-estimation contrast of contrast in the propensity to report a Correct value between HIGH and LOW treatments

CorrectCalculations	χ^2	p
0	0.8	0.3706
1	0.33	0.5646
2	0.04	0.8408
3	0.07	0.7862
4	0.79	0.3751
5	2.69	0.101
6	5.78	0.0162
7	8.98	0.0027
8	11.1	0.0009
9	11.66	0.0006
10	10.65	0.0011
11	8.3	0.004
12	5.29	0.0215
13	2.65	0.1036
14	0.99	0.3195
15	0.22	0.6357
16	0	0.9465
17	0.06	0.8055
18	0.24	0.6252
19	0.46	0.4997
20	0.66	0.4153

Table B.6: Post-estimation contrast of contrast in the propensity to report a PartialFalse value between HIGH and LOW treatments

Appendix C

For Chapter 3

C.1 Experimental Materials

C.1.1 Experiment - Student Instructions

Instructions

Welcome to the experiment.

This is an experiment into economic decision making. Please read the instructions carefully, as your payment will be a function of your decisions. There are no right or wrong decisions.

You will be paid a base fee of £3 for taking part in this experiment and may have the opportunity to earn a bonus fee through your choices in the experiment.

This experiment is in three parts. In the first part you will be asked a series of questions by the computer for which you will need to make a number of selections.

The second part of the experiment will present you with a task that you must complete to qualify for the third section of the experiment. The details of the second part of the experiment will be given after you have completed the questions in the first part. The instructions for the third part of the experiment will be presented to you only after you have qualified for the third section.

When you are ready to begin, please click on the button below to start the first part of the experiment.

Question
Please let us know your gender
Please let us know your age
Please let us know your course of study

C.1.2 Experiment Part 1: Task 1

Note: The third question (Please let us know your course of study) was replaced by the question Please let us know your nationality for the Amazon Mechanical Turk sample

Below are five pairs of descriptions. Circle one point on each scale to indicate how much you think each description sounds like you. For example:

If a pair of descriptions describe you equally well, then mark the centre of the scale

If you are slightly more like description 1 than description 2, then mark the scale slightly closer to description 1

If description 2 is exactly right and description 1 is not like you at all, then mark the scale right next to description 2

C.1.3 Experiment Part 1: Task 2

Note: The £ sign was replaced by the $ sign for the Amazon Mechanical Turk sample

Please consider the following hypothetical investment.

This investment has a 50/50 chance of success. If successful, any amount invested will return 2.5 times the value invested. However, if not successful then it will return nothing. For example, if all £10 were to be invested, then if it is successful, it will return £25, but if unsuccessful it will return £0.

You may invest any portion of the £10 (from 0 to 10), and keep the remainder.

To fix ideas, consider the following hypothetical examples:

Suppose you invested £3, and kept the remaining £7.

If the investment is successful, then you will receive £7.50 from the investment,

Description 1	Description 2
Someone who is talkative, outgoing, is comfortable around people, but could be bossy and attention seeking	Someone who is a reserved, private person, doesn't like to draw attention to themselves and can be shy around strangers
Someone who is a reserved, private person, doesn't like to draw attention to themselves and can be shy around strangers	Someone who is generally trusting and forgiving, is interested in people, but can be taken for granted and finds it difficult to say no
Someone who is sensitive and excitable, and can be tense	Someone who is relaxed, unemotional rarely gets irritated and seldom feels blue
Someone who likes to plan things, likes to tidy up, pays attention to details, but can be rigid or inflexible	Someone who doesn't necessarily work to a schedule, tends to be flexible, but disorganised and often forgets to put things back in their proper place
Someone who is a practical person who is not interested in abstract ideas, prefers work that is routine and has few artistic interests	Someone who spends time reflecting on things, has an active imagination and likes to think up new ways of doing things, but may lack pragmatism

plus £7 that you kept for a total of £14.50. However, if the investment is not successful, then you will receive nothing from the investment and only retain the amount you kept for a total of £7.

Suppose you invested £8 and kept the remaining £2.

If the investment is successful, then you will receive £20 from the investment, plus £2 that you kept for a total of £22. However, if the investment is not successful, then you will receive nothing from the investment and only retain the amount you kept for a total of £2.

How much of the £10 would you invest?

Option 1	£100.00 in thirteen months' time
Option 2	£101.50 in thirteen months' time
Option 3	£103.00 in thirteen months' time
Option 4	£104.50 in thirteen months' time
Option 5	£106.00 in thirteen months' time
Option 6	None of the above

C.1.4 Experiment Part 1: Task 3

Note: The £ sign was replaced by the $ sign for the Amazon Mechanical Turk sample

In this section you are asked to choose from a list of hypothetical options corresponding to different amounts of money to be paid in the future. What is the lowest amount you would prefer to have instead of £100 in one month's time?

So if the first option you would prefer is £104.50 in thirteen months' time to £100 in one month's time, (such that you would not have preferred £100, £101.50 or £103.00 in thirteen months' time), you would select option 4.

Please select which option you would choose.

C.1.5 Experiment Part 2 Instructions

Experiment Part 2 - Instructions

In this second part of the experiment you will engage in a computerised task. The task requires that you re-position sliders that will be presented on the screen. There will be 48 sliders, each of which will initially be positioned at the left hand side of the bar, at position 0. Your task is to position as many of these sliders as possible in the centre of the bar, at position 50, in 100 seconds.

The image below shows how the screen will look once the experiment has started with the slider in the top left hand corner having been moved to the centre position.

We will conduct three practice rounds to allow you to gain familiarity and some practice with the task before moving onto a final round that will be used to determine if you qualify for the final part of the experiment. You must re-position 15 of the 48 sliders to the centre of the bar the final round of this task to qualify for Part 3 of the experiment.

C.1.6 Experiment Part 3 Instructions

Experiment Part 3 - Instructions

In this, final, part of the experiment you may have the opportunity to increase your payment from today's experiment.

The computer will perform a lottery by drawing a random whole number from 1 to 100, with all numbers having the same probability of being selected, that is the chance of the value 12 being drawn is 1-in-100, as is the chance of the value 77 being drawn.

You will be randomly partnered with another subject in the experimental session who has also qualified for part 3 of the experiment to form a pair. The other person in your pair has a lottery value of their own, which is independent of the the lottery value you receive. Your task in this part of the experiment is to report a value of the lottery to the person you have been paired with. The other person will be presented with the same set of instructions and asked to make the same choices.

Once the experimental session is over, one of the members of your pair will be randomly chosen by the computer to be a type A, whereas the other will then be classed as a type B. The computer will calculate your payment based on the type you are assigned to and the appropriate choices made.

For the person selected to be **type A**, the payment will be 5 pence times the value they report

For the person selected to be **type B**, the payment will be 5 pence times the value the type A person they were paired with reports

The following table presents 4 examples of outcomes. In each case the lottery produces the same values for the two players (You and your paired other). The differences arise from the values chosen and the random selection of who is

	You	Other	You	Other	You	Other	You	Other
Value Drawn	30	60	30	60	30	60	30	60
Report if A	30	60	30	60	80	20	80	20
Report if B	30	60	30	60	20	70	20	70
Type Assigned	A	B	B	A	A	B	B	A
Payment	1.50	1.50	3.00	3.00	4.00	4.000	1.00	1.00

assigned to type A and who is assigned to type B.

Only the value you choose to report for the type you are selected to be will be shown to the experimental participant you are paired with. Only the corresponding value chosen by the subject you are paired with will be shown to you. The selections not used and the actual values drawn in the lottery will not be revealed.

As you will not know what type you will be allocated before you make your choices, we will ask you to make the choice as to what message you would send to the other person in your pair if you were of type A.

Your assigned type, payment and the value reported by the participant you were paired with will be available online after the experiment has finished. These details will also be sent to you in an e-mail.

Payment will be made through the student expenses claim system. We will request that the appropriate payment is made to you. To ensure that you receive this payment, you will need to register the appropriate bank account details in the student record system. failure to register the correct bank details in the student record system will lead to a delay in your payment. Further details on how to ensure your bank details are registered will be included in the e-mail that will detail your payment.

C.1.7 Experiment Part 3 Quiz

Experiment Part 3 - Quiz

We will now ask a number of simple questions to check your understanding of the experiment. These questions are shown below. Once you have correctly answered the questions below, you will be shown your value for the lottery and asked to make your decisions.

Number	Question
1	Will you be able to know the value of lottery draw for the person you are paired with?
2	How many choices will you be required to make in this part of the experiment?
3	If you are selected to be a type B person, how many times the value of your lottery draw that you report will you be paid?
4	If you are selected to be a type A person, how many times the value of the lottery draw you report will the person you are paired with be paid?
5	What is the maximum payment available for this part (part 3) of the experiment?
6	What is the minimum payment available for this part (part 3) of the experiment?

C.1.8 Experiment Part 3 - Lottery Choice

Your value for the lottery draw is **12**. Please indicate the values you would like to report to experimental participant you will be paired with in the case where you are selected to be type A.

In the case where I am assigned to type **A**, I will report a value of:

C.1.9 Experiment Part 3 - Payment

Experiment - Complete

Thank you completing this experiment. You have now been randomly matched with another participant from this experimental session. You have been assigned to type A.

The value you declared as a type A was 100.

Your payment for today's experiment will therefore be £8.00.

Please ensure that your bank account details in the Student Record System are correct. Failure to enter correct details into SRS will delay any payment due to you. You can access the student record system through exhub or directly through SRS.

C.1.10 Experiment - Mechanical Turk Workers Recruitment

We are conducting an academic study about individual decision making. We would like to invite you to take part in our study, in which you will have to make a number of choices, some of which may give you the opportunity to make additional earnings from participating.

This study should take 30 minutes to complete.

You will be paid $3 for doing this study.

You may earn more depending on the decisions that you make during the experiment. Any extra amount you earn will be paid as a bonus.

If you wish to complete the study, please copy and past the link below into a new web browser window or tab. At the end of the study, you will receive a code to paste into the box below to receive your payment.

Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box.

C.1.11 Experiment - Mechanical Turk Workers Pre-instruction Page

The purpose of the study

The purpose of this study is to understand individual decisions. The study is being conducted by Prof. Miguel A. Fonseca, and Shaun Grimshaw

Participation and withdrawal

<p>Participation in this study is completely voluntary and you are free to withdraw from this study at any time without prejudice or penalty. If you wish to withdraw, simply stop completing the task and close your browser. If you do withdraw from the study, the materials that you have completed to that point will be deleted and will not be included in the study.

What is involved?

We will present you with a series of decisions which may carry financial consequences to you and, depending on the decision, to someone else. Participation in this study will take approximately 15 minutes and you will be paid $1 for taking part, and possibly more depending on your decisions.

Risks

Participation in this study should involve no physical or mental discomfort, and no risks beyond those of everyday living. If, however, you should find any question or procedure to be invasive, you are free to omit answering or participating in that aspect of the study.

Confidentiality and security of data

All data collected in this study will be stored confidentially. All data is anonymous, and cannot be linked to you. The data you provide will only be used for the specific research purposes of this study.

Ethics Clearance and Contacts

This study has been cleared in accordance with the ethical review processes of the University of Exeter Business School. You are free to discuss your participation with project staff (m.a.fonseca@exeter.ac.uk ; sbg203@exeter.ac.uk). Alternatively, you may leave a message with the Research Team at the University of Exeter (business-school-research-office@exeter.ac.uk). By proceeding with the experiment, you provide consent that you understand the information provided and agree to participate in this study. You also understand that your participation is voluntary and you may withdraw at any time.

Please click next below for further instructions about the experiment.

C.1.12 Experiment - Mechanical Turk Workers Instructions

Instructions

Welcome to the experiment.

This is an experiment into economic decision making. Please read the instructions carefully, as your payment will be a function of your decisions. There are no right or wrong decisions.

You will be paid a base fee of $3 for taking part in this experiment and may have the opportunity to earn a bonus fee through your choices in the experiment. At the end of the experiment you will be given a code to enter into mechanical turk to claim your base fee.

This experiment is in three parts. In the first part you will be asked a series of questions by the computer for which you will need to make a number of selections.

The second part of the experiment will present you with a task that you must complete to qualify for the third section of the experiment. The details of the second part of the experiment will be given after you have completed the questions in the first part. The instructions for the third part of the experiment will be presented to you only after you have qualified for the third section.

When you are ready to begin, please click on the button below to start the first part of the experiment.

C.2 Regression Results

	UpLie			DownLie		
DV: classifup	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
LotteryDraw	-0.0177***	-0.0208***	-0.0202***	0.01	0.0123	0.0151
	(0.006)	(0.0062)	(0.0062)	(0.0087)	(0.0086)	(0.0098)
SKEW	-0.2181	-0.3459	-0.3012	0.0326	0.0968	-0.1504
	(0.4314)	(0.4422)	(0.4548)	(0.7294)	(0.7287)	(0.7837)
SKEW	-0.0008	0.0017	-0.0007	0.0014	0.0002	0.0041
	(0.0087)	(0.009)	(0.0093)	(0.0124)	(0.0124)	(0.0136)
STUDENT	1.4118*	1.366	1.3722	-5.8507**	-4.691**	-5.2133**
	(0.8535)	(0.8552)	(0.8567)	(2.5237)	(2.0523)	(2.0524)
STUDENT LotteryDraw	-0.0178	-0.0188	-0.0199	0.0702**	0.0572**	0.0768***
	(0.0143)	(0.0143)	(0.0135)	(0.0316)	(0.0266)	(0.0249)
SKEW STUDENT	-0.1334	-0.01	-0.1911	6.1262**	4.88**	6.0073***
	(1.0735)	(1.0819)	(1.0611)	(2.7436)	(2.3025)	(2.3278)
SKEW STUDENT LotteryDraw	0.0046	0.002	0.0058	-0.0825**	-0.0655**	-0.0655**
	(0.0197)	(0.0195)	(0.0192)	(0.036)	(0.0317)	(0.0292)
Error		-0.7898***	-0.7887***		1.3108**	1.0144*
		(0.2634)	(0.2819)		(0.5771)	(0.5248)
Age			-0.0007			0.0192
			(0.0127)			(0.0216)
Male			0.3238			0.0724
			(0.2365)			(0.3572)
Extroversion			0.0056			-0.2028**
			(0.0639)			(0.0981)
Agreeableness			-0.014			-0.0057
			(0.0662)			(0.1017)
Emotional Stability			-0.0053			-0.1372
			(0.0649)			(0.094)
Conscientiousness			0.0648			0.0776
			(0.0699)			(0.1073)
Openness			-0.0711			-0.0293
			(0.07)			(0.1116)
Allocation			0.0663			-0.0635
			(0.0469)			(0.0814)
Slifer			0.0057			-0.0429*
			(0.0142)			(0.0242)
Prelottery			-0.0011**			0.0001
			(0.0005)			(0.0005)
Lotteryinstructions			0.0007			-0.0191***
			(0.001)			(0.0049)
Lotteryquiz			-0.0034***			0.0015
			(0.0013)			(0.0022)
Lotterydecision			0.0108**			0.0078
			(0.0047)			(0.0074)
cons	0.9842***	1.7437***	1.9759*	-1.5863***	-2.8731***	1.3766
	(0.3273)	(0.4268)	(1.0351)	(0.5484)	(0.7778)	(1.4827)
N	461	461	461			
Pseudo R2	0.07	0.10	0.20			
LL	-418.8	-407.0	-359.2			

[1]

Table C.1: Multinomial logistic estimates of the classification of subject choices into UpLie, Honest and DownLie

[1]Model 1 based on 3.3. Model 2 incorporates the term Error. Model 3 incorporates a number of subject responses and observed times for steps within the experiment

LotteryDraw	BASE		SKEW	
	χ^2	p	χ^2	p
0	6.56	0.0104	6.2	0.0128
10	6.99	0.0082	7.09	0.0077
20	7.18	0.0074	7.7	0.0055
30	7.04	0.008	7.35	0.0067
40	6.32	0.0119	5.56	0.0183
50	4.61	0.0317	3.29	0.0697
60	2.31	0.1289	1.67	0.1958
70	0.65	0.4208	0.8	0.37
80	0.01	0.911	0.38	0.5393
90	0.43	0.5104	0.17	0.6796
100	2.27	0.132	0.07	0.7898

Table C.2: Post-estimation contrast of the propensity to report a UpLie between student and AMT subjects

LotteryDraw	BASE		SKEW	
	χ^2	p	χ^2	p
0	4.09	0.043	5.27	0.0217
10	3.85	0.0497	5.54	0.0186
20	3.34	0.0676	5.37	0.0205
30	2.57	0.1092	4.46	0.0346
40	1.56	0.2117	2.83	0.0926
50	0.54	0.4613	1.21	0.2717
60	0.03	0.873	0.29	0.5927
70	0.05	0.8163	0.01	0.9357
80	0.04	0.8419	0.07	0.7932
90	0.09	0.7698	0.27	0.6048
100	0.83	0.3625	0.5	0.4817

Table C.3: Post-estimation contrast of the propensity to report a Honest value between student and AMT subjects

LotteryDraw	BASE		SKEW	
	χ^2	p	χ^2	p
0	4.36	0.0368	0.76	0.3836
10	6.04	0.014	0.94	0.3313
20	8.61	0.0033	1.19	0.2763
30	12.47	0.0004	1.47	0.2254
40	17.15	0	1.73	0.1889
50	18.5	0	1.84	0.175
60	11.62	0.0007	1.76	0.1841
70	3.53	0.0602	1.58	0.2083
80	0.15	0.7007	1.4	0.2375
90	0.51	0.477	1.25	0.2644
100	1.98	0.1594	1.14	0.2858

Table C.4: Post-estimation contrast of the propensity to report a DownLie between student and AMT subjects

	FullLie			PoisitvePartialLie			DownLie		
DV: classif	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model1	Model 2	Model 3
LotteryDraw	-0.0091	-0.0136*	-0.014**	-0.0327***	-0.0337***	-0.0325***	0.0102	0.0121	0.0145
	(0.0067)	(0.0072)	(0.0071)	(0.0085)	(0.0086)	(0.0092)	(0.0088)	(0.0087)	(0.0099)
SKEW	-0.1928	-0.407	-0.3922	-0.3948	-0.4344	-0.3704	0.0415	0.1067	-0.1305
	(0.5038)	(0.5246)	(0.5379)	(0.5059)	(0.507)	(0.5428)	(0.7351)	(0.7332)	(0.7895)
SKEW	-0.0077	-0.004	-0.0059	0.0128	0.0135	0.0108	0.0013	0.0003	0.0043
	(0.0101)	(0.0104)	(0.0104)	(0.0116)	(0.0116)	(0.0127)	(0.0125)	(0.0124)	(0.0137)
STUDENT	1.3891	1.2935	1.1166	1.2968	1.281	1.64*	-5.8528**	-4.6895**	-5.1509**
	(0.9377)	(0.942)	(0.9532)	(0.9395)	(0.9374)	(0.9952)	(2.5287)	(2.0569)	(2.0983)
STUDENT LotteryDraw	-0.0221	-0.0236	-0.0225	-0.0071	-0.0075	-0.0111	0.0702**	0.0574**	0.0765***
	(0.0162)	(0.0164)	(0.0154)	(0.0174)	(0.0173)	(0.0173)	(0.0317)	(0.0267)	(0.0255)
SKEW STUDENT	-0.7981	-0.6187	-0.4797	0.4419	0.4773	0.084	6.1295**	4.9017**	6.0241**
	(1.2537)	(1.2827)	(1.2593)	(1.1716)	(1.1718)	(1.1767)	(2.7536)	(2.3133)	(2.375)
SKEW STUDENT LotteryDraw	0.0218	0.0182	0.0164	-0.0171	-0.0177	-0.0086	-0.0825**	-0.066**	-0.0654**
	(0.0232)	(0.0232)	(0.0223)	(0.0241)	(0.024)	(0.0248)	(0.0362)	(0.0319)	(0.0297)
Error		-1.1582***	-1.0701***		-0.2964	-0.4047		1.2971**	1.0137**
		(0.2907)	(0.3045)		(0.3144)	(0.3465)		(0.5754)	(0.5204)
Age			-0.0004			0.0028			0.0198
			(0.0161)			(0.0152)			(0.0216)
Male			0.4452			0.1979			0.0598
			(0.2878)			(0.2821)			(0.358)
Extroversion			0.0036			0.0127			-0.2042**
			(0.0752)			(0.0773)			(0.0986)
Agreeableness			-0.0339			-0.0099			-0.0077
			(0.0774)			(0.0825)			(0.1013)
Emotional Stability			0.0359			-0.044			-0.1386
			(0.076)			(0.0771)			(0.0941)
Conscientiousness			0.0611			0.0603			0.0785
			(0.0821)			(0.0838)			(0.1075)
Openness			-0.1177			-0.0264			-0.0305
			(0.0813)			(0.0819)			(0.1125)
Allocation			0.0772			0.0528			-0.0645
			(0.0581)			(0.0554)			(0.0812)
Slider			0.0236			-0.016			-0.0444*
			(0.018)			(0.0173)			(0.0243)
Prelottery			-0.0007			-0.0017***			0
			(0.0005)			(0.0006)			(0.0005)
Lotteryinstructions			0.0022*			-0.0019			-0.0196***
			(0.0012)			(0.0015)			(0.0049)
Lotteryquiz			-0.0036**			-0.0032*			0.0014
			(0.0017)			(0.0017)			(0.0022)
Lotterydecision			-0.0004			0.0189***			0.0097
			(0.0072)			(0.0059)			(0.0081)
cons	0.0849	1.1682**	0.5639	0.6231	0.9116*	2.4155*	-1.5954***	-2.8624***	1.5271
	(0.3765)	(0.4861)	(1.2557)	(0.388)	(0.4845)	(1.2739)	(0.5558)	(0.7783)	(1.4845)
N	461	461	461						
Psudeo R2	0.07	0.10	0.20						
LL	-568.1	-522.5	-490.9						

Table C.5: Multinomial logistic estimates of the classification of subject choices into FullLie, PositivePartialLie, Honest and DownLie[2]

[2]Model 1 based on 3.3. Model 2 incorporates the term Error. Model 3 incorporates a number of subject responses and observed times for steps within the experiment

LotteryDraw	STUDENT		AMT	
	χ^2	p	χ^2	p
0	1.37	0.2414	0	0.955
10	1.52	0.2176	0.16	0.6907
20	1.58	0.2093	0.93	0.3349
30	1.39	0.239	2.51	0.1134
40	0.91	0.3392	4.45	0.0349
50	0.41	0.522	5.74	0.0166
60	0.1	0.7511	5.91	0.015
70	0	0.9874	5.37	0.0205
80	0.08	0.7755	4.61	0.0318
90	0.36	0.5501	3.88	0.0489
100	0.72	0.3963	3.26	0.0712

Table C.6: Post-estimation contrast of the propensity to report a FullFalse value between BASE and SKEW treatments

LotteryDraw	BASE		SKEW	
	χ^2	p	χ^2	p
0	0.41	0.5197	7.6	0.0058
10	1.05	0.3049	7.76	0.0053
20	2.22	0.1365	7.12	0.0076
30	3.74	0.0531	4.8	0.0284
40	4.65	0.0311	2.14	0.1439
50	4.2	0.0405	0.68	0.4091
60	3.03	0.0819	0.14	0.7062
70	1.9	0.1681	0	0.9497
80	1.05	0.3067	0.03	0.866
90	0.41	0.5199	0.12	0.728
100	0.03	0.8601	0.24	0.6257

Table C.7: Post-estimation contrast of the propensity to report a PositivePartial-False value between student and AMT subjects

Task	STUDENT	AMT
Instructions	0.976	0.894
Personality	0.521	0.55
Investment	0.515	0.704
Time	0.539	0.833
Sliderinstructions	0.388	0.983
Slider	0.915	0.718
Lotteryinstructions	0.703	0.309
Lotteryquiz	0.779	0.208
Lotterydecision	0.936	0.176
Prelottery	0.24	0.909
Total	0.674	0.49

Table C.8: Kolmogorov-Smirnov test p-values of distributions between the BASE and SKEW treatments

Task	z	p
Lotterydecision	2.615	0.0089
Lotteryquiz	-3.157	0.0016
Lotteryinstructions	7.897	< 0.001
Slider	-4.039	< 0.001
Sliderinstructions	4.052	< 0.001
Time	9.956	< 0.001
Investment	4.898	< 0.001
Personality	6.753	< 0.001
Instructions	3.475	0.0005
Pretime	4.657	< 0.001
Total	4.41	< 0.001

Table C.9: Mann-Whitney tests of Median time differences between student and AMT subjects

Bibliography

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. Journal of Public Economics, 113 (6), 96–104.

Abeler, J., Nosenzo, D., and Raymond, C. (2016). “Preferences for truth-telling”. Working Paper, CEDEX Discussion Paper Series 2016-13.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. Economics Letters, 80, 123–129.

Akerlof, G. A. (1983). Loyalty filters. American Economic Review, 73 (1), 54–63.

Akerlof, G. A. and Kranton, R. E. (2002). Economics and Identity. The Quarterly Journal of Economics, 115 (3), 715–753.

Akiaya, E., Hanaki, N., and Ishikawa, R. (2017). In is not just confusion! Strategic uncertainty in an experimental asset market. The Economic Journal, 127 (605), F563–F580.

Allingham, M. G. and Sandmo, A. (1972). Income Tax Evasion: A Theoretical Analysis. Journal of Public Economics, 1, 323–338.

Alm, J., Bloomquist, K. M., and McKee, M. (2015). On the external validity of Laboratory tax compliance experiments. Economic Inquiry, 53 (2), 1170–1186.

Alm, J. and McKee, M. (2004). Tax Compliance as a Coordination Game. Journal of Economic Behaviour and Organization, 54 (3), 297–312.

Andersen, S., Harrison, G. W., Lauc, M. I., and Rutström, E. E. (2010). Preference heterogeneity in experiments: Comparing the field and laboratory. Journal of Economic Behaviour and Organization, 73 (2), 209–224.

Anderson, J., Burks, S. V., Carpenter, J., Gotte, L., Maurer, K., Nosenzo, D., Potter, R., Rocha, K., and Rustichini, A. (2013). Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. Experimental Economics, 16, 170–189.

Andreoni, J. (1995). Coperation in public-goods experiments: Kidness or confusion? American Economic Review, 85 (4), 891–904.

Andreoni, J., Erard, B., and Freinsein, J. (1988). Tax Compliance. Journal of Economic Literature, 36, 818–860.

Arrow, K. J. (1970). "Political and Economic Evaltion of Social Effects and Externalities". In: *The Analysis of Public Output*. NBER. Chap. 1, pp. 1–30.

Baron, J. and Ritov, I. (1994). Reference Points and Omission Bias. Organizational Behavior and Human Decision Processes, 59, 475–498.

Bartneck, C., Duenser, A., Moltchanova, E., and Zawieska, K. (2015). Comparing the similarity of responses received from studies in amazon's mechanical turk to studies conducted online and with direct recruitment. PLoS ONE, 10, e0121595.

Becker, G. (1968). Crime and Punishment: An Economic Approach. Journal of Political Economy, 76, 169–217.

Behnk, S., Barreda-Tarrazona, I., and García-Gallego, A. (2014). The role of ex post transparency in information transmission—An experiment. Journal of Economic Behavior and Organization, 101, 45–64.

Bellemare, C. and Kröger, S. (2007). On representative social capital. European Economic Review, 51 (1), 183–202.

Bellemare, C., Kröger, S., and Van Soest, A. (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. Econometrica, 76 (4), 815–839.

Belot, M., Dutch, R., and Miller, L. (2015). A comprehensive comparison of students and non-studentsin classic experimental games. Journal of Economic Behavior and Organization, 113, 26–33.

Bernheim, B. D. (1988). "Financial Illiteracy, Education, and Retirement Saving". In: *Living with Defined Contribution Pensions*.

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2013). Simplification and Saving. Journal of Economic Behavior and Organization, 95, 130–145.

Bigoni, M., Camera, G., and Casari, M. (2013). Strategies of cooperation and punishment among students and clerical workers. Journal of Economic Behaviour and Organization, 94, 172–182.

Biziou-van-Pol, L., Haenen, J., Novaro, A., Liberman, A. O., and Capraro, V. (2015). Does telling white lies signal pro-social preferences? Judgment and Decision Making, 10 (6), 538–548.

Bloomquist, K., Emblom, E., Johns, D., and Langtieg, P. (2012). Estimates of the Tax Year 2006 Individual Income Tax Underreporting Gap. IRS-TPC Research Conference, Publication 1500, 69–78.

Bortolotti, S., Casari, M., and Pancotto, F. (2015). Norms Of Punishment: Experiments with students and the general population. Economic Inquiry, 53 (2), 1207–1223.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., and Satorra, A. (2002). One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments. American Economic Review, 92 (5), 1687–1701.

Brown, C. L. and Krishna, A. (2004). The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice. Journal of Consumer Research, 31, 529–539.

Bruner, D., Jones, M., McKee, M., and Vossler, C. A. (2015). “Tax Reporting Behavior: Underreporting Opportunities and Pre-populated Tax Returns”. Mimeo.

Burks, S., Carpenter, J., Gotte, L., and Rustichini, A. (2009). Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment. Proceedings of the National Academy of Sciences of the United States of America, 106 (19), 7745–7750.

Cai, H. and Wang, J. T.-Y. (2006). Overcommunication in strategic information transmission games. Games and Economic Behavior, 56 (1), 7–36.

Campbell, E. Q. (1964). The Internalization of Moral Norms. Sociometry, 27 (4), 391–412.

Cappelen, A. W., Nygaard, K., Sørensen, E. Ø., and Tungodden, B. (2015). Comparison of Students and a Representative Population. The Scandinavian Journal of Economics, 117 (4), 1306–1326.

Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2013). When do we lie? Journal of Economic Behavior and Organization, 93, 258–265.

Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. Experimental Economics, 11, 282–298.

Chakravarty, S., Ma, Y., and Maximiano, S. (2011). *Lying and Friendship*. Working Papers 1008. Purdue University, Department of Consumer Sciences.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. Journal of Economic Behavior and Organization, 81, 1–8.

— (2013). Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. Journal of Economic Behavior and Organization, 91, 93–100.

Chesher, A. and Irish, M. (1987). Residual analysis in the grouped and censoted normal linear model. Journal of Econometrics, 34 (1), 33–61.

Childs, J. (2012). Gender differences in lying. Economics Letters, 114 (2), 147–149.

Choo, C. Y. L., Fonseca, M. A., and Myles, G. D. (2016). Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment. Journal of Economic Behavior and Organization, 124, 102–114.

Cialdini Robert B. Kallgren, C. A. and Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. Advances in Experimental Social Psychology, 24, 201–234.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. (2006). Managing Social Norms for Persuasive Impact. Social Influence, 1 (1), 3–15.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. Journal of Personality and Social Psychology, 58, 1015–1026.

Cleave, B. L., Nikiforakis, N., and Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. Experimental Economics, 16 (3), 372–382.

Cohn, A., Fehr, E., and Marechal, M. A. (2014). Business culture and dishonesty in the banking industry. Nature, 516 (7529), 86–89.

Conrads, J. (2014). *The Effect of Communication Channels on Lying*. Cologne Graduate School Working Paper Series 05-06. Cologne Graduate School in Management, Economics and Social Sciences.

Crump, M. J. C., McDonnell, J. V., and Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. Plos ONE, 8 (1), e57140.

Cushman, F., Young, L., and Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment. Psychological Science, 17 (12), 1082–1089.

De Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schelthammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). The Neural Basis of Altruistic Punishment. Science, 305 (5688), 1254–1258.

Deetlefs, J., Chylinski, M., and Ortmann, A. (2015). MTurk 'Unscrubbed': Exploring the Good, the 'Super', and the Unreliable on Amazon's Mechanical Turk. Working Paper, UNSW Business School Research Paper 2015-20A.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in Everyday Life. Journal of Personality and Social Psycholog, 70 (5), 979–995.

Depositario, D. P. T., Nayga, R. M. J., Wu, X., and Laude, T. P. (2009). Should students be used as subjects in experimental auctions? Economics Letters, 102, 122–124.

Dolan, P., Hallsworth, M., Halpern, D., King, D., and Vlaev, I. (2010). "MINDSPACE: Influencing Behaviour Through Public Policy". Institute of Government, London, UK.

Dreber, A. and Johannesson, M. (2008). Gender differences in deception. Economics Letters, 99, 197–199.

Duncan, D. and Li, D. (2017). Liar Liar: Experimental Evidence of the Effect of Confirmation-Reports on Dishonesty. Southern Economic Journal. In Press.

Erat, S. and Gneezy, U. (2012). White Lies. Management Science, 58 (4), 723–733.

European Commission (2012). Improving Tax Governance in EU Member States: Criteria for Successful Policies. URL: http://ec.europa.eu/economy_finance/publications/occasional_paper/2012/pdf/ocp114_en.pdf. [Accessed Jun 23, 2015].

Exadaktylos, F., Espın, A., and Branas-Garza, P. (2013). Experimental subjects are not different. Scientific Reports, 3 (1213). DOI: DOI:10.1038/srep01213.

Falk, A., Meier, S., and Zehnder, C. (2013). Do Lab Experiments Misrepresent Social Preferences? The Case of Self-Selected Student Samples. Journal of the European Economic Association, 11 (4), 839–852.

Financial Times (2012). HP takes $8.8bn hit over Autonomy. URL: https://www.ft.com/content/6d2dd40e-3311-11e2-aabc-00144feabdc0. [Accessed 20 Feb 2017].

— (2014). Tesco in turmoil after profits overstatement. URL: https://www.ft.com/content/5823a7cc-4279-11e4-9818-00144feabdc0. [Accessed 20 Feb 2017].

Financial Times (2017). VW admits guilt and pays $4.3bn emissions scandal penalty. URL: https://www.ft.com/content/d998b804-d81a-11e6-944b-e7eb37a6aa8e. [Accessed 20 Feb 2017].

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in Disguise: An Experimental Study on Cheating. Journal of the European Economic Association, 11 (3), 525–547.

Fonseca, M. A. and Grimshaw, S. B. (2017). Do behavioral nudges in pre-populated tax forms affect compliance? Experimental evidence with real taxpayers. Journal of Public Policy and Marketing. In Press.

Frechette, G. R. (2015). “Laboratory Experiments: Professionals versus Students”. In: *Handbook of Experimental Economic Methodology*. Ed. by G. R. Frechette and A. Schotter. UK: Oxford University Press, pp. 360–390.

— (2016). “Experimental Economics Across Subject Populations”. In: *The Handbook of Experimental Economics, vol 2*. Ed. by A. E. Kagel John H. abd Roth. USA: Princeton University Press, pp. 435–480.

Gächter, S., Herrmann, B., and Thöni, C. (2004). Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. Journal of Economic Behavior and Organization, 55, 505–531.

Gale, W. G. and Holtzblatt, J. (1997). On the Possibility of a No-return Tax System. National Tax Journal, 50 (3), 475–485.

Gaudecker, H.-M. von, Soest, A. van, and Wengström, E. (2012). Experts in experiments: How selection matters for estimated distributions of risk preferences. Journal of Risk and Uncertainty, 45 (2), 159–190.

Gibson, R., Tanner, C., and Wagner, A. F. (2013). Preferences for Truthfulness: Heterogeneity among and within Individuals. The American Economic Review, 103 (1), 532–548.

Gigerenzer, G. (2010). Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality. Topics in Cognitive Science, 2, 528–554.

Gill, D. and Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. American Economic Review, 102 (1), 469–503.

Gneezy, U. (2005). Deception: The Role of Consequences. The American Economic Review, 95 (1), 384–394.

Gneezy, U., Kajackaite, A., and Sobel, J. (2016). “Lying Aversion and the Size of the Lie”. Working Paper.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. Quarterly Journal of Economics, 112 (2), 631–646.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. Journal of Economic Behavior and Organization, 93, 293–300.

Goffman, E. and Best, J. (2005). *Interaction Ritual: Essays in Face to Face Behavior*. Aldine Transaction, New Brunswick.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. Journal of Consumer Research, 25 (3), 472–482.

Government Accountability Office (GAO) (2005). Highlights of a GAO Forum: The Federal Government's Role in Improving Financial Literacy (GAO-05-93SP). URL: `http://www.gao.gov/cgibin/getrpt?GAO-05093SP`. [Accessed Jun 23, 2015].

Greenberg, E., Smeets, P., and Zhurakhocska, L. (2015). "Promoting truthful communication through ex-post disclosure". Working Paper, UCSD.

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. Economics Letters, 107 (2), 291–296.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. Journal of the Economic Science Association, 1 (1), 114–125.

Gylfason, H. F., Arnardottir, A. A., and Kristinsson, K. (2013). More on gender differences in lying. Economics Letters, 119 (1), 94–96.

Hao, L. and Houser, D. (2017). Perceptions, Intentions and Cheating. Journal of Economic Behaviour and Organization, 133, 52–73.

Harrison, G. W. and List, J. A. (2004). Field Experiments. Jounral of Economic Literature, 42 (4), 1009–1055.

Healy, P. M. and Palepu, K. G. (2003). The fall of Enron. Journal of Economic Perspectives, 17 (2), 3–26.

Hilbig, B. E. and Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. Journal of Research in Personality, 57, 72–88.

HMRC (2015a). Making Tax Easier: The End of the Tax Return. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/413975/making-tax-easier.pdf. [Accessed Jun 23, 2015].

— (2015b). Measuring tax gaps 2015 edition. URL: https://www.gov.uk/government/uploads/system/uploads/attachment%5C_data/file/470540/HMRC-measuring-tax-gaps-2015-1.pdf. [Accessed 20 Feb 2017].

Holm Håkan, J. and Kawagoe, T. (2010). Face-to-face lying – An experimental study in Sweden and Japan. Journal of Economic Psychology, 31 (3), 310–321.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: conducting experiemnts in a real labour market. Experimental Economics, 14 (3), 399–425.

Horton, J. J. and Chilton, L. B. (2010). "The Labor Economics of Paid Crowdsourcing". In: *Proceedings of the 11th ACM Conference on Electronic Commerce*. EC '10. Cambridge, Massachusetts, USA: ACM, pp. 209–218. ISBN: 978-1-60558-822-3.

Houser, D. and Kurzban, R. (2002). Revisiting kindness and confusion in public-goods experiments. American Economic Review, 92 (4), 1062–1069.

Hurkens, S. and Kartik, N. (2009). Would I lie to you? On social preferences and lying aversion. Experimental Economics, 12 (2), 180–192.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. Journal of Economic Behavior and Organization, 93 (0), 328–336.

Johnson, E. J. and Goldstein, D. G. (2003). Do defaults save lives? Science, 302, 1338–1339.

Johnson, E. J., Hershey, J., Meszaros, J., and Kunreuther, H. (1993). Framing, Probability Distortions, and Insurance Decisions. Journal of Risk and Uncertainty, 7, 35–51.

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., and Weber, E. U. (2012). Beyond Nudges: Tools of a Choice Architecture. Marketing Letters, 23, 487–504.

Kanuk, L. and Berenson, C. (1975). Mail surveys and response rates: A literature review. Journal of Marketing Research, 12 (4), 440–453.

Kees, J., Berry, C., Burton, S., and Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. Journal of Advertisting, 47 (1), 141–155.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2012). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. Econometrica, 79 (3), 651–692.

Knutson, B., Adams, C. M., Fong, G. M., and Hommer, D. (2001). Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens. Journal of Neuroscience, 21 (16), RC159–RC159.

Kotakorpi, K. and Laamanen, J.-P. (2015). “Complexity, Salience and Income Tax Reporting Behaviour: Evidence from a Natural Experiment”. Mimeo.

Lamberton, C. (2013). A spoonful of choice: How allocation increases satisfaction with tax payments. Journal of Public Policy and Marketing, 32 (2), 223–238.

López-Pérez, R. and Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. Experimental Economics, 16 (3), 233–247.

Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. Journal of Economic Behavior and Organization, 70 (1-2), 81–92.

Lusardi, A. and Mitchell, O. (2007). Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education. Business Economics, 42 (1), 35–44.

Lusardi, A. and Mitchell, O. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. Journal of Economic Literature, 52 (1), 5–44.

Madrian, B. C. and Shea, D. F. (2001). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. Quarterly Journal of Economics, 116 (4), 1149–1187.

Mazar, N., Amir, O., and Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. Journal of Marketing Research, 45, 633–644.

Mazar, N. and Ariely, D. (2006). Dishonesty in Everyday Life and Its Policy Implications. Journal of Public Policy and Marketing, 25 (1), 117–126.

Mazar, N. and Hawkins, S. A. (2015). Choice Architecture in Conflicts of Interest: Defaults as Physical and Psychological Barriers to (Dis)honesty. Journal of Experimental Social Psychology, 59, 113–117.

Milgrom, P. and Shannon, C. (1994). Monotone Comparative Statics. Econometrica, 62 (1), 157–180.

National Audit Office (2015). Tackling Tax Fraud: How HMRC Responds to Tax Evasion, the Hidden Economy and Criminal Attacks. URL: https://www.nao.org.uk/wp-content/uploads/2015/12/Tackling-tax-fraud-how-HMRC-responds-to-tax-evasion-the-hidden-economy-and-criminal-attacks.pdf. [Accessed Jun 23, 2015].

Nerlove, M. and Press, S. J. (1973). Univariate and Multivariate Log-Linear and Logisitc Models. RAND-R1306-EDA/NIH, Santa Monica.

New York Times (2014). Facebook Apologizes for Overstating Video Metrics. URL: https://www.nytimes.com/2016/09/24/business/media/facebook-apologizes-for-overstating-video-metrics.html. [Accessed 20 Feb 2017].

O'Doherty, J. P., Deichmann, R., Critchley, H. D., and Dolan, R. J. (2002). Neural Responses During Anticipation of a Primary Taste Reward. Neuron, 33 (5), 815–826.

OECD (2006). Using Third Party Information Reports to Assist Taxpayers Meet their Return Filing Obligations— Country Experiences With the Use of Pre-populated Personal Tax Returns. URL: http://www.oecd.org/tax/administration/36280368.pdf. [Accessed Jun 23, 2015].

Onu, D. and Oats, L. (2014). "Social norms and tax compliance". Discussion paper 006-14, Tax Administration Research Centre, University of Exeter.

Pagan, A. and Vella, F. (1989). Diagnostic test for models based on individual data: A survey. Jounrnal of Applied Econmetrics, 4, S29–S59.

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

Peer, E., Vosgerau, J., and Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behavior Research Methods, 46 (4), 1023–1031.

Peeters, R., Vorsatz, M., and Walzl, M. (2015). Beliefs and truth-telling: A laboratory experiment. Journal of Economic Behavior and Organization, 113, 1–12.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K., and Dickert, S. (2006). Numeracy and Decision Making. Psychological Science, 17 (5), 407–413.

Piovesan, M. and Wengström, E. (2009). Fast or fair? A study of response times. Economics Letters, 105, 193–196.

Reeson, A. and Dunstall, S. (2009). "Behavioural Economics and Complex Decision-Making: Implications for the Australian Tax and Transfer System". CSIRO/CMIS Report No. 09/110.

Reno, R. R., Cialdini, R. B., and Kallgren, C. A. (1993). The Transsituational Influence of Social Norms. Journal of Personality and Social Psychology, 64 (1), 104–112.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A Neural Basis for Social Cooperation. Neuron, 35 (2), 395–405.

Rode, J. (2010). Truth and trust in communication: Experiments on the effect of a competitive context. Games and Economic Behavior, 68 (1), 325–338.

Rubinstein, A. (2013). Response time and decision making: An experimental study. Judgment and Decision Making, 8 (5), 540–551.

Ruffle, B. J. and Tobol, Y. (2017). Clever enough to tell the truth. Experimental Economics, 20 (1), 130–155.

Samuelson, W. and Zeckhauser, R. (1988). Status Quo Bias in Decision Making. Journal of Risk and Uncertainty, 1 (1), 7–59.

Sánchez-Páges, S. and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. Games and Economic Behavior, 61 (1), 86–112.

— (2009). Enjoy the Silence: An Experiment on Truth-telling. Experimental Economics, 12, 220–241.

Schultz, W. P., Nolan, J. M., Cialdini, R. B., J., G. N., and Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. Psychological Science, 18 (5), 429=434.

Shalvi, S., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011a). Ethical Manoeuvring: Why People Avoid Both Major and Minor Lies. British Journal of Management, 22, S16–S27.

Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011b). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. Organizational Behavior and Human Decision Processes, 115 (2), 181–190.

Sherif, M. and Sherif, C. W. (1953). *Groups in harmony and tension*.

Slonim, R., Wang, C., Garbarinoc, E., and Merrett, D. (2013). Opting-in: Participation bias in economic experiments. Journal of Economic Behavior and Organization, 90, 43–70.

Smith, N. C., Goldstein, D. G., and Johnson, E. J. (2013). Choice Without Awareness: Ethical and Policy Implications of Defaults. Journal of Public Policy and Marketing, 32 (2), 159–172.

Spiliopoulos, L. and Ortmann, A. (2017). The BCD of response time analysis in experimental economics. Experimental Economics. In Print, 1–51.

Spranca, M., Minsk, E., and Baron, J. (1991). Omission and Commission in Judgment and Choice. Journal of Experimental Social Psychology, 27, 76–105.

Sussman, A. B. and Olivola, C. Y. (2011). Axe the tax: Taxes are disliked more than equivalent costs. Journal of Marketing Research, 48, S91–S101.

Talwar, V., Murphy, S. M., and Lee, K. (2007). White lie-telling in children for politeness purposes. International Journal of Behavioral Development, 31 (1), 1–11.

Teper, R. and Inzlicht, M. (2011). Active transgressions and moral elusions: Action framing influences moral behavior. Social Psychological and Personality Science, 2 (3), 284–288.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*.

The Guardian (2017). Snapchat accused of lying about user numbers to inflate value of IPO. URL: https://www.theguardian.com/technology/2017/jan/05/snapchat-accused-lying-user-numbers-ipo-investors. [Accessed 20 Feb 2017].

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica, 26 (1), 24–36.

Utikal, V. and Fischbacher, U. (2013). Disadvantageous lies in individual decisions. Journal of Economic Behavior and Organization, 85, 108–111.

Vanberg, C. (2017). Who never tells a lie? Experimental Economics, 20 (2), 448–459.

Vendemia, J. M. C., Buzan, R. F., and Simon-Dack, S. L. (2005). Reaction Time of Motor Responses in Two-Stimulus Paradigms Involving Deception and Congruity with Varying Levels of Difficulty. Behavioural Neurology, 16 (1), 25–36.

Vrij, A. and Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. Psychology, Crime and Law, 5 (3), 203–215.

Wang, J. T., Spezio, M., and Camerer, C. F. (2010). Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. American Economic Review, 100 (3), 984–1007.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. Stata Journal, 12 (2), 308–331.

Woods, S. A. and Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. European Journal of Personality, 19 (5), 373–390.

Wright, P. (2002). Marketplace Metacognition and Social Intelligence. Journal of Consumer Research, 28, 677–682.

Yu, J. and Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. Journal of Marketing Research, 20 (1), 36–41.