Three experiments exploring how preferences, motivations and incentives influence behaviour

Author: Zack Dorner

Supervisors: Professor Lata Gangadharan,

Associate Professor Paul Raschky & Doctor Anke Leoux

A thesis submitted for the degree of Doctor of Philosophy at

Monash University in November 2017

in the Department of Economics

Copyright notice

Abstract

Preferences and motivations drive behaviours. Particular underlying preferences, such as risk preferences or pro-environmental preferences, can be found to influence the decisions people make about complex goods. On the other hand, changing incentives through imposing extrinsic incentives or through technology change can also impact behaviours. The aim of this thesis is to contribute to the body of knowledge on how preferences and motivations drive behaviour, while also exploring how altering incentives can change behaviours in expected or unexpected ways. The research projects in this thesis are applied, specifically within environment and health settings, with a key element of interest being heterogeneity. I utilise an experimental economics methodology throughout, as it provides a powerful means of investigating hypotheses derived from theory, within controlled decision contexts that have real consequences for subjects. To this end, this thesis is comprised of three main papers, all of which utilise fully incentivised laboratory experiments, either as lab-in-the-field experiments (the first and second papers) or as a pure lab experiment (the third paper).

In the first paper, my co-authors and I run a fully incentivised risk preferences experiment alongside a stated preference survey to model utility over intrinsic risk. Participants' estimated coefficients of constant relative risk aversion (CRRA) are incorporated into preference estimation for new sources of municipal water supply to test the hypotheses that supply risk (vulnerability to drought) and new technology risk are important intrinsic attributes. Controlling for water quality and cost, we find that supply risk – and not technology risk – is a partial determinant of participants' choices. The second paper investigates intrinsic motivation, how it is affected by a range of extrinsic incentives and whether it has a role in health outcomes. We find only the low power monetary incentive raises effort when imposed, whereas the high power monetary incentive crowds out effort after removal. However, the high power monetary incentive raises effort for low motivation individuals and does not significantly crowd out their effort once removed. Intrinsic motivation is found to partially explain waist-to-height ratio. For the third paper, I investigate whether there is a behavioural rebound effect, expressed as a reduction in pro-environmental effort after an improvement in the environmental efficiency of technology. I also test for moral licensing, where individuals who endogenously choose an energy efficient product subsequently give themselves a psychological licence to reduce their level of pro-environmental effort further. I find evidence for a behavioural rebound effect, which is estimated to be 32% in a laboratory setting. Moral licensing also occurs, increasing the size of the behavioural rebound effect, and is strongest among subjects with a higher level of pro-environmental orientation of their attitudes and beliefs.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

In the case of chapters 2 and 3, the work came from active collaboration with co-authors. My contribution to the work involved the following:

Thesis chapter	Title	Nature and % of my contribution	Co-author name(s); nature and % of co-authors contribution
2	Preferences for Intrinsically Risky Attributes	Data analysis, write up and copy editing (50%).	Daniel A. Brent: Input into data analysis, input into manuscript (20%). Anke Leroux: Research question, experimental design and coordination of data collection, input into manuscript (30%).
3	Intrinsic motivation, health outcomes and the crowding out effect of temporary extrinsic incentives: A lab-in-the-field experiment	Input into research question, input into experimental design, data collection, data analysis, write up and copy editing (65%).	Emily Lancsar: Research question, experimental design, input into data analysis, input into manuscript (35%).

Signature: [redacted]

Name: Zack Dorner

Date: 24 November 2017

Acknowledgements

As I am sure is true with the many who have gone before me, and the many to come, writing this thesis has involved highs and lows. I could not have done it without the help of numerous people along the way. To those who I name below, and all the others I have unintentionally left off, I thank you very much.

The first acknowledgement must go to my amazing wife, Raven Cretney. We have been on this journey together from our undergraduate degrees, through to you inevitably beating me to getting a PhD. It has been so important to be able to share this experience with you, as well as to have someone to get out of the city with when we have needed a break. I am looking forward to all that we will achieve together in the future.

I have been blessed to have a great team of supervisors at Monash. To Lata Gangadharan, Paul Raschky and Anke Leroux, thank you for all the time, comments and support you have provided me. I have learned a lot from each of you and it really has been an academic apprenticeship to have been able to work with you three.

To my dad, Daniel, my mum, Mary, my sister, Rebecca, and my large and wonderful extended family around the world. You have all helped encourage and instill in me a love of problem solving, as well as a creative and stubborn streak (more so the latter than the former in my case!). I could not have done a PhD without the many things you have taught me. It has also been great to have had my in laws just "down the road" while Raven and I have been worked on our theses. Michelle, Mike, Summer, Alex and Luke – the many meals, great conversations and fun activities we shared were more important to completing this thesis than you might realise!

There have been a great many people with whom I have done everything from discussing ideas, to working at a more detailed level. Particular thanks of course to my co-authors, Danny Brent, Anke Leroux and Emily Lancsar. You have truly been integral to this thesis. Then there are those I have had discussions with or have commented on my work. They include the great teachers we benefited from in our first year of the PhD programme. I found the Department of Economics at Monash a fantastically supportive place where everyone was prepared to give the PhD students a generous amount of time. With apologies in advance for inevitably forgetting someone, I thank Robert Johnston, John Rolfe, Liang Choon Wang, Michael Ward, Tim Cason, Kenan Kalayci, Philip Grossman, Elias Khalil, Birendra Rai, Klaus Abbink, Frans de Vries, John Gibson, Jeffrey LaFrance, Andreas Lange, Frank Scrimgeour, Steven Tucker, Bruce Preston, Siew Ling Yew, plus the many people who have sat through and commented on presentations I have given, including a number of seminars at Monash, plus the Environmental Economics Workshop in 2015 and 2016, AARES 2016 and 2017, AAEA 2016, the 29th PhD Conference in Economics and Business in 2016, and NZARES 2017.

The department at Monash also attracts a high quality of PhD students – quality colleagues as well as quality friends. I am glad to have been able to share the tough assignments and research anxieties with those in my cohort, as well as time to relax and unwind. I also benefited from the wisdom of the "older" students, and the fresh energy from the new faces each year. In no particular order, and with apologies for those I have inevitably left off, I acknowledge Alex, Michael, Sakiba, Son, Sabine, Joe, Kevin, Klaus, Vilas, Anwesha, Ashani, Vy, Leo and honorary Monash PhD student, Peter.

Finally, my thesis has also been made possible by, and greatly benefited from a range of funding. First, the Cooperative Research Centre for Water Sensitive Cities (CRCWSC) was outstanding to be involved with as a PhD student. I learnt a lot of useful skills from the various programmes run for PhD students, and it was fantastic to be able to meet so many fellow students from myriad disciplines. I am also grateful for the funding I received from the Department of Economics at Monash. Lastly, this research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Contents

List of tables

List of figures

Chapter 1

Introduction

How people come to a decision, and why they behave in certain ways, are large and very deep questions. Shedding light on these questions can help us better understand ourselves, why we do what we do, and how we might make better decisions in the future. The importance of these questions is demonstrated by how they permeate everything from the early works by Adam Smith (1759; 1776), to more modern influential books such as *Nudge* by Richard Thaler and Cass Sunstein (2009). These pieces consider not just how internal human motivations affect behaviours, but also how external forces can impact behaviours and internal motivations themselves. In writing this thesis it has been my endeavour to push the frontier of human knowledge out in some very small way on how preferences, motivations and incentives influence behaviour.

While understanding our behaviours better is an intrinsically valuable task, there are many other good reasons for doing so. For our health, adding up all the little things that we do can, over time, lead to a great improvement, or, conversely, a large reduction in our quality of life. The complex society that we live within can also greatly benefit or suffer from all the decisions that we make regarding how we use scarce resources, including both the resources we have within society, as well as the resources available to us from the natural world. Undertaking applied research with the potential for impact within these important contexts is the second major motivation for this thesis.

In the first paper presented in this thesis, my co-authors and I develop a method to better understand individuals' choices in a stated preference experiment, by enhancing our model of their choices with data from an incentivised risk preferences experiment. Incentivised experiments are commonly viewed as substitutes for, rather than complements to, stated preference methods. While the former are founded in revealed behaviour, the latter are able to characterise preferences in situations that cannot be directly observed. We leverage the distinct strengths of each approach to model preferences in a situation where the utility derived from a risky attribute of a good is determined by one's tolerance for risk.

We combine a fully incentivised risk experiment in the field with a stated preference discrete choice experiment (DCE) to model utility for intrinsic risk. A door-to-door survey of 981 participants in a drought-prone areas (Melbourne and Sydney) elicits preferences for alternative sources of municipal water, conditional on water price and quality. Participants' estimated coefficients of constant relative risk aversion (CRRA) are incorporated into preference estimation to test the hypotheses that supply risk (vulnerability to drought) and new technology risk are important intrinsic attributes for new water sources. Controlling for water quality and cost, we find that supply risk – and not technology risk – is an important determinant of participants' choices.

In the second paper, my co-author and I consider the role of intrinsic motivation in health behaviours, and how extrinsic incentives can reduce intrinsic motivation. The impetus behind the paper is the substantial international policy interest in incentivising healthy behaviours. When considering incentives, particularly monetary incentives, policymakers should be mindful of the potential for the crowding out of intrinsic motivation, where effort may be reduced when incentives are applied, and after they are removed. Furthermore, with regards to monetary incentives, there is conflicting evidence on whether it is better to go big (Gneezy and Rustichini, 2000b) or to go small (Pokorny, 2008).

In the paper, we investigate the effect on intrinsic motivation of a range of extrinsic incentives, which vary by size and type, both during their application and after their removal. Additionally, we investigate whether intrinsic motivation predicts health outcomes. The laboratory experiment is comprised of a rich within and between subject design that allows us to estimate a differences-in-differences model of the treatment effects. Our subject pool is a heterogeneous adult population. Subjects are given four time limited rounds of a real effort task. Round 1 measures intrinsic motivation. Extrinsic incentives are applied for round 2, varying between subjects. Extrinsic incentives are removed for rounds 3 and 4 to measure crowding out and persistence of crowding out. On average, we find support for the "pay – but do not pay too much" rule (of Pokorny, 2008). However, we find that "pay enough or don't pay at all" (of Gneezy and Rustichini, 2000b) better fits the results for low motivation individuals. The high power monetary incentive is most likely to crowd out intrinsic motivation after its removal. Intrinsic motivation is found to partially explain waist-to-height ratio, demonstrating the relevance of our findings for health policy, and an example of how motivations measurably drive behaviour.

The third and final paper builds on the first two by investigating whether underlying pro-environmental preferences might lead to a behavioural rebound effect, using a real effort laboratory experiment. Significant attention has been paid to leveraging behavioural motivators (non-price interventions) to increase energy conservation (Allcott and Mullainathan, 2010). Technological change that improves energy efficiency is also important (Global Commission on the Economy and Climate, 2014). While both focus on reducing energy use, these two strands of literature have yet to be joined to consider what behavioural effects might result from technology change. The direct rebound effect is the increase in consumption due an increase in energy efficiency and can be modelled as the rational response to a change in relative prices (Chan and Gillingham, 2015). This paper investigates whether there might also be a behavioural rebound effect by looking at two potential sources.

The first potential source of a behavioural rebound effect is where pro-environmental behaviours are reduced after an improvement in energy efficiency. Second, moral licensing may increase the behavioural rebound effect if individuals who buy an energy efficient product subsequently give themselves psychological licence to reduce their pro-environmental behaviours even further. I develop a novel laboratory experiment to investigate these mechanisms. They can be cleanly isolated in the laboratory without the many confounds potentially present in the field, such as other motivations to reduce energy usage like saving money. Subjects must decide how to allocate their effort, in a real effort task, between earning money for themselves and reducing damages to a tree planting charity. I find evidence for a behavioural rebound effect, which is estimated to be 32% in this laboratory setting. Moral licensing also occurs, increasing the size of the behavioural rebound effect, and it is strongest among subjects with a higher level of pro-environmental orientation of their attitudes and beliefs. The main driver of pro-environmental effort is shown to be beliefs about social norms. This paper extends the core model of the rebound effect, and the findings can help inform policies to encourage pro-environmental behaviours within the context of constantly improving environmental efficiency of technology.

As covered above, I employ an experimental economics methodology for all three papers in this thesis. I undertake artefactual field experiments (or lab-in-the-field experiments) for the first two papers.[1] The third paper utilises a laboratory experiment. In all three papers, I take the approach of measuring some variable in a laboratory style task, and applying it to an outcome in the field. For the first paper, it is risk preferences, which are used to help understand decisions made in a DCE. In the second paper, I measure intrinsic motivation and relate it to health outcomes. In the third paper, the laboratory experiment I develop measures the behavioural rebound effect and moral licensing behaviour, with implications for how we can expect pro-environmental behaviours in the field to evolve as the environmental efficiency of technology improves.

An experimental economics methodology provides the researcher with a high degree of control over the treatment conditions provided to participants, or subjects, of the research. This control provides precision in testing hypotheses and theories, for example by providing information on risk preferences from a fully incentivised task that thus had real consequences for subjects, as in my first paper. Another example from my third paper is allowing for credibly exogenous changes to energy efficiency. In the field the decision to upgrade to a more energy efficient technology is endogenous. There are

[1] Using the nomenclature of Harrison and List (2004) and Viceisza (2016), respectively.

of course limitations to the method, as with any method, including that subjects may behave differently when they know they are being observed (Levitt and List, 2007). Both the advantages and limitations are discussed in each chapter, particularly where they are relevant to the research aims and hypothesis testing. They are addressed in terms of experimental design and interpretation of results, with the aim of maximising the potential from the use of the experimental methods applied within.

The first paper of this thesis is presented in the following chapter. The second and third papers proceed in Chapters 3 and 4. Final remarks and additional areas for research are presented in Chapter 5. The bibliography for all chapters in the thesis, and appendices for each individual chapter, are found at the end of the document.

Chapter 2

Preferences for intrinsically risky attributes

2.1 Introduction

Experimentally elicited preferences are widely utilised to predict behaviour in the field (Fehr and Leibbrandt, 2011; Cavalcanti et al., 2013; Gneezy et al., 2016). A key strength of fully incentivised experiments is that preference elicitation is founded in revealed behaviour; in contrast, stated preference methods are able to characterise preferences in situations that cannot be directly observed. Thus, while there are opportunities for combining revealed and stated methods (Adamowicz et al., 1994; Whitehead et al., 2008), incentivised experiments are more likely to be seen as substitutes rather than complements to stated preference methods. For example, consumer preferences for food are elicited using either stated choice methods (Scarpa et al., 2012; Meas et al., 2015) or experiments in the laboratory and in the field (Melton et al., 1996; Lusk and Coble, 2005). In other instances, incentivised experiments are used to validate the results of stated preference methods (List and Shogren, 1998). In this paper we leverage the distinct strengths of each approach and use information on respondents' attitudes from an incentivised lab-in-the-field experiment to augment the estimates in a stated choice study, thereby gaining additional insights about the respondents' preferences for intrinsic attributes that would otherwise remain hidden.

Our approach relates to Lancaster's (1966) theory of consumption, which states that utility is derived not from the good or service itself, but rather from its characteristics or attributes. Building on this premise, stated choice methods make predictions about changes in utility over alternatives that result from changes in their attributes. While the analyst has control over the extrinsic attributes for each alternative presented, specific alternatives may also have intrinsic attributes. One can think of intrinsic attributes as the residual attributes that are left unspecified in a stated choice experiment. Consider a travel mode choice experiment that offers the choice between public transit and automobile travel with extrinsic attributes for the travel time, reliability, and cost. The unspecified intrinsic attributes for public transit may be inconvenience, the ability to read while commuting, and warm glow from making an environmentally friendly choice. In the empirical analysis of the choice experiment these intrinsic attributes are generally bundled into an alternative specific constant (ASC) that communicates the aggregate preferences for public transit relative to driving, conditional on the extrinsic attributes.

In some settings, however, it may be desirable to assess individuals' preferences or beliefs for an attribute without explicitly defining it as an extrinsic attribute. For example, the risk of an accident can be presented as an attribute in the travel choice example, but it would not necessarily capture the respondents' pre-existing beliefs about the risk of cars relative to transit, which are formed by idiosyncratic information unobservable to the analyst. Moreover, preferences for varying degrees of travel risk depend on the respondents' attitudes to risk that are similarly unobservable. Failing to allow for the respondents' perceptions of, and preferences for, intrinsic attributes can be problematic. An important example of this is when the propensity to participate in a survey depends on risk attitudes in a systematic way. In this paper we show that leveraging information about risk attitudes to model preferences for unspecified intrinsic risk attributes improves the model fit and yields significantly different estimates of marginal utilities.

Our application combines a fully incentivised risk experiment with a stated preference approach. The risk experiment, involving incentivised decisions over binary monetary lotteries (similar to Holt and Laury, 2002) is randomly allocated to a subsample of 981 households that participate in a door-to-door survey, where respondents are asked to choose among six alternative sources of water to augment their city's central water supply. The survey uses a discrete choice experimental design (DCE), where alternative water sources vary with respect to allowed water use and cost to the household. The survey is conducted in Melbourne and Sydney, Australia, where residents frequently experience droughts that result in restrictions to household water use as well as controversial public investments to boost central water supply.[1] Therefore, public knowledge about centralised sources of water provision is high, making it likely that consumers have well-formulated beliefs regarding the intrinsic risks of different supply sources.

Ex ante we hypothesise that there are two sources of intrinsic risk affecting the choices made by participants. These sources of risk are intentionally *not* mentioned in the information materials provided to the participants of the DCE to ensure participants are not biased towards responding to these risks more than they would otherwise. First, some sources (a new dam, stormwater harvesting and interbasin transfer pipeline) are dependent on weather and therefore may not provide sufficient water security during periods of drought. We term this risk 'supply risk'. Additionally, certain sources (stormwater harvesting and recycled water) provide water via new and somewhat unproven technologies, which may be of concern to some consumers. We label this intrinsic attribute 'technology risk'. We argue that Australian households have well-formed perceptions of these two risks based on the extensive public discourse surrounding water supply augmentation during the Millennium Drought. For example, Dolnicar et al. (2014) show that only 28% of respondents believe that the current, reservoir-sourced tap water can save Australia from drought, whereas they are much more confident about the ability of desalination (77%) and recycled water (83%) to sustain water supplies during a drought. While 90% of Australian respondents believe their current water is safe to drink, only 54% think this is true of recycled water, which is, according to 73% of respondents, also prone to technological failure. These findings by Dolnicar et al. (2014) motivate our hypotheses that

[1] There is an extensive literature on the acceptance of various forms of water supply in Australia, see Fielding et al. (2015) and the papers cited for more information.

supply and technology risks may be important determinants of preferences for different water sources.

The paper is organised as follows. The next section positions this study within the revealed and stated preference literature on risk and risk attitudes. The theoretical framework is outlined in Section 2.3, followed by a brief description of the experimental design and summary statistics. Section 2.5 summarises the empirical framework, Section 2.6 describes the main results and Section 2.7 concludes.

2.2 Risk in Preference Elicitation

Agricultural and environmental policies tend to have strong elements of risk and uncertainty regarding outcomes (Pindyck, 2007), and a recent focus of the DCE literature has been on improving the methodology to deal with outcome-related risk. For example, Glenk and Colombo (2013) add risk of failure as an extrinsic attribute for policy options aimed at increasing soil carbon in Scotland, and hence reducing greenhouse gas emissions. They use this data to estimate the preferences of their participants with regards to the level of uncertainty of policies and find the non-linear expected utility theory model performs best. Other DCE studies are concerned with outcome-related risk surrounding the level of environmental quality of a particular lake (Roberts et al., 2008), policies to improve fish numbers and size in popular angler spots (Wielgus et al., 2009) and policies to improve the environmental quality in the Great Barrier Reef (Rolfe and Windle, 2015). These studies demonstrate that the addition of an extrinsic attribute that captures outcome related risk alters the stated preferences compared with studies that do not explicitly allow for outcome related risks (Roberts et al., 2008; Wielgus et al., 2009). Our results complement these findings in that we also find an effect on estimated preferences when allowing for intrinsic risk attributes in a stated choice setting. Moreover, we find that this effect varies systematically with the respondent's risk attitude.

Similarly, there is a growing literature that focuses on risk attitudes within the contexts of flood insurance (Botzen and Van Den Bergh, 2012; Botzen and van den Bergh, 2012; Petrolia et al., 2013), investments in energy efficiency (Qiu et al., 2014), wildfire protection

(Bartczak et al., 2015) and reducing health risks (Lusk and Coble, 2005; Anderson and Mellor, 2008; Cameron and DeShazo, 2013; Andersson et al., 2016). Botzen and Van Den Bergh (2012) analyse the role of increased flood risk from climate change on the market for flood insurance. They investigate how consumers respond to low-probability risks and changes in risk, as well the role of communicating risk probabilities in risk-related decisions. In a revealed preference setting Petrolia et al. (2013) elicit risk attitudes in order to investigate the role of risk aversion on flood insurance uptake. In most of these settings risk has a direct effect on the preferences for the good and is explicitly modeled as an attribute in a choice experiment (Botzen and Van Den Bergh, 2012; Botzen and van den Bergh, 2012), or as a driver of private purchase decisions (Petrolia et al., 2013). In our specific setting where risk is an intrinsic characteristic of the good we also find that risk attitudes matter to consumer choices.

Where risk is a central feature of the good, such as the probability of a flood for flood insurance, it can be modeled explicitly. However, in settings such as the deployment of a new technology, where risk perceptions about the new technology are complex, it may be preferable to consider risk as an intrinsic attribute and allow respondents to communicate risk preferences through their choices. For example, self reported data reveals that risk averse people are less likely to purchase energy efficient appliances (Qiu et al., 2014) and take longer to adopt new farming technologies (Liu, 2013). Other examples relate to "range anxiety" for electric cars, where consumers face an increase in the risk of being stranded from choosing an electric car over a petrol version (Hidrue et al., 2011). The analyst cannot credibly decouple these risks as extrinsic attributes, and it is this type of intrinsic risk that is the focus of this paper.

The research that is closest to our own from a methodological perspective is Newell and Siikamäki (2014) and Newell and Siikamäki (2015). Those studies experimentally elicit individual discount rates to help assess if respondents in a stated choice experiment on buying a new hot water system trade off between upfront and operating costs in a cost efficient manner. In contrast, our focus is on eliciting preferences for intrinsic attributes by leveraging information on risk preferences. Importantly, our approach can be generalised to link existing preferences to a wide range of intrinsic attributes, thereby helping to improve the estimation of stated preference models. For example, conditional cooperation elicited in public goods games can be linked to the intrinsic attributes of public transit and car pooling versus driving alone in a travel mode choice experiment.

2.3 Theoretical Framework

We begin with a random utility model (McFadden, 1973) of householders' choices over a set of J alternative municipal water sources. Utility U of individual i from choosing water source j for choice occasion t is given by

$$U_{ijt} = V_{ijt} + \epsilon_{ijt}, \tag{2.1}$$

where V_{ijt} is a linear function of the observable source attributes, allowed use (quality level) and cost per kL consumed, and ϵ_{ijt} is a random component incorporating all other factors that may affect U_{ijt}. In particular, if V_{ijt} contains ASCs, these dummies incorporate attributes that are intrinsic to the water source such as supply or technology risks. Individual i chooses water source j for choice t when:

$$U_{ijt} \geq U_{ikt} \quad \forall j, k \in J, j \neq k. \tag{2.2}$$

A standard empirical application of this model assumes the observable component, V_{ijt}, to be linear and additively separable in its elements. Thus, in our base model:

$$V = \boldsymbol{\beta}_j \mathbf{X}_j + \boldsymbol{\beta}_q \mathbf{X}_q + \beta_c C, \tag{2.3}$$

where $\boldsymbol{\beta}_j$ is a vector of the ASCs for each water source $\mathbf{X}_j$, relative to the source that is represented by the omitted categorical dummy. The vector of coefficients $\boldsymbol{\beta}_q$ is associated with the different levels of allowed use, $\mathbf{X}_q$, and β_c is the coefficient on cost per kL of water consumed.

In addition to our base model we propose an alternative model specification that explicitly allows for heterogeneous risk attitudes toward a subset of water sources that may be perceived as intrinsically risky. In particular, a subset of sources may be perceived as risky if their supply depends on exogenous factors such as rainfall or if the technology that is used to provide water is new and unproven. From the outset, we are agnostic about which type of risk may be important and test models where a dummy variable X_r describes different types of risk. As before, it is assumed that, independently of allowed use and cost, each water source provides some utility that is certain from the respondents' perspective. This component enters the utility function in the standard linear form, $\beta_j X_j$. An additional utility component is linked to the perceived riskiness of particular sources. Because of its intrinsic nature, the risk-related component of utility only enters the utility function through an interaction with risk-preferences. Therefore, in most studies that do not estimate risk preferences, this component of utility is not observable. The importance of risk attitudes for explaining heterogeneous preferences is our central hypothesis of interest.

Retaining the additively-separable specification of equation (2.3) the risk-related utility component is accommodated as follows,

$$V = \boldsymbol{\beta}_j \mathbf{X}_j + \boldsymbol{\beta}_q \mathbf{X}_q + \beta_c C + \beta_{r,h} f\left(X_{r,h}, \gamma_i\right), \tag{2.4}$$

where the sign of $\beta_{r,h}$ indicates whether the participants perceive source(s) $h \subset J$ as risky. The magnitude of $\beta_{r,h}$ represents the weight of this intrinsic risk on utility. $X_{r,h}$ is the risk variable that takes on the value 2 if the source(s) is affected by risk relative to all sources assigned a value of 1. The parameter γ_i denotes each individual's constant relative risk aversion (CRRA) in the non-linear specification $f\left(X_{r,h}, \gamma_i\right) = \left(\frac{X_{r,h}^{1-\gamma_i} - 1}{1-\gamma_i}\right)$.[2] This parameter is estimated independently using an incentivised lab-in-the field risk experiment. Thus, the utility that is attributable to $X_{r,h}$ depends on each individual's CRRA parameter. A risk loving individual is characterised by $\gamma_i < 0$, a risk neutral individual by $\gamma_i = 0$ and a risk averse individual has $\gamma_i > 0$.[3]

[2]The definition of the risk variable $X_{r,h} \in \{1, 2\}$ in conjunction with the CRRA functional form ensures that the risk-related component of utility is 0 when $X_{r,h} = 1$, while varying continuously in the degree of risk aversion for $X_r = 2$.

[3]This specification assumes that, given observed risk attitudes, intrinsic risk-related utility can be

Table 2.1: Interpreting the coefficient on $\beta_{r,h}$ for source $j \in h$ relative to source $k \notin h$.

	Risk Loving $\gamma_i < 0$	Risk Neutral $\gamma_i = 0$	Risk Averse $\gamma_i > 0$	Perception of Source j
$\beta r, h > 0$	$U_j > U_k$	$U_j = U_k$	$U_j < U_k$	Relatively Risky
$\beta r, h < 0$	$U_j < U_k$	$U_j = U_k$	$U_j > U_k$	Relatively Safe

This assumes that the all other non-risk related attributes are for sources j and k are equal such as the ASC, cost, and quality. This follows the notation in equation (2.5) where $X_{r,j} = 2$ and $X_{r,k} = 1$.

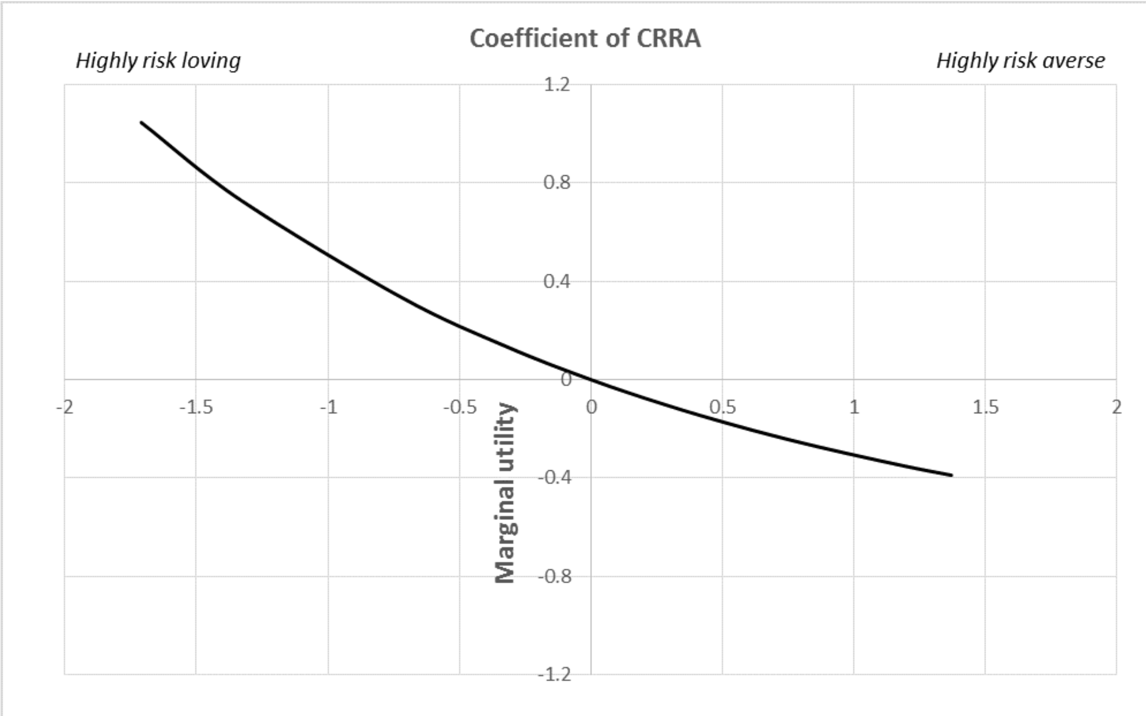
To illustrate the differences between the base model and risk-augmented model we compare the marginal utility implied by each model from choosing a water source j relative to source k with the same level of quality and cost. In the base model without risk preferences, the marginal utility of choosing source j over source k is $\beta_j - \beta_k$, which is the utility derived from the ASC for water source j. In the extended model, the marginal utility from choosing source j over source k takes into account both utility components: the deterministic change in utility, β_j, as well as the change in utility that is due to the relative riskiness of each source and is described by the non-linear combination of $\beta_{r,j}$ and γ_i. Assuming $X_{r,j} = 2$ and $X_{r,k} = 1$, the marginal utility of choosing source j over source k is given by

$$U_{ij} - U_{ik} = \beta_j + \beta_{r,j}\left(\frac{2^{1-\gamma_i} - 1}{1 - \gamma_i}\right) - \beta_k - \beta_{r,k}\left(\frac{1^{1-\gamma_i} - 1}{1 - \gamma_i}\right) \tag{2.5a}$$
$$= (\beta_j - \beta_k) + \beta_{r,j}\left(\frac{2^{1-\gamma_i} - 1}{1 - \gamma_i}\right). \tag{2.5b}$$

The sign of $\beta_{r,j}$ contains information about the relative riskiness of the two sources as perceived by the respondents. Table 2.1 shows how the sign of $\beta_{r,j}$ interacts with risk aversion parameter to impact utility, assuming equality of all non-risk related attributes. Importantly, Table 2.1 shows how the sign of $\beta_{r,j}$ yields information about how respondents perceive the riskiness of source j relative to source k. This allows us to test for intrinsic risk preferences for various water supply sources.

fully separated out from the ASCs. For example, it assumes the utility from the supply risk of a water source can be captured separately from the utility of choosing a particular water source by the term $\beta_{r,h} f(X_{r,h}, \gamma_i)$, where $X_{r,h}$ is supply risk.

Figure 2.1: Marginal utility of switching from source k to source j where source j is considered riskier than source k and $\beta_{rj} > 0$.



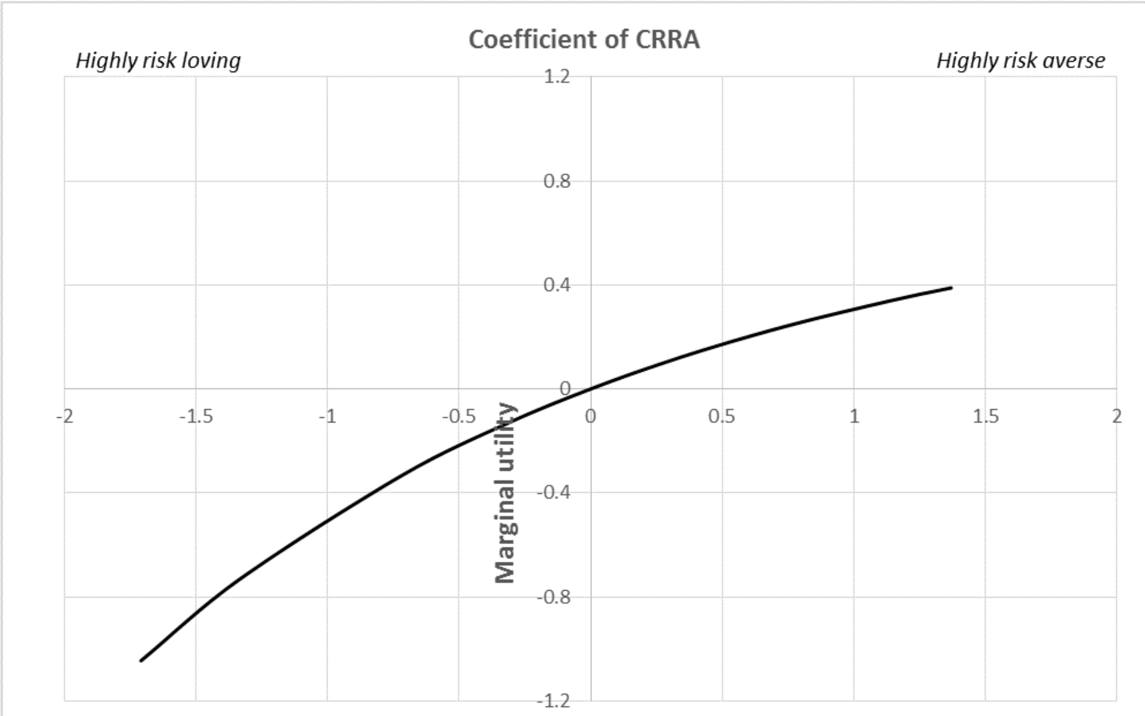
To help clarify how risk is incorporated into our model we graphically illustrate the marginal utility for switching from source k to source j for different levels of risk aversion given the value of $\beta r, j$.[4] Figure 2.1 illustrates that when $\beta_{r,j}$ is positive the water source j is perceived as *riskier* than source k: a switch from source k to this riskier, but otherwise equally preferred, source j brings positive utility to risk loving individuals and negative utility to risk averse individuals.[5] In contrast, a negative coefficient ($\beta_{r,j} < 0$) in equation (2.5) indicates that source j is perceived to be *safer* than source k, so that the switch from source k to the safer, but otherwise equal, source j brings negative utility to risk loving individuals and positive utility to risk averse individuals. The marginal utility as a function of risk aversion when $\beta_{r,j} < 0$ is shown in Figure 2.2. As seen in both Figures 2.1 and 2.2 the risk component of utility is 0 for a risk neutral consumer ($\gamma_i = 0$).

Whether a particular intrinsic attribute that is common to a subset of sources is considered risky by participants, and thus given a significant weight in determining their

[4]Similar to Table 2.1, Figures 2.1 and 2.2 follows the notation in equation (2.5) where $X_{r,j} = 2$ and $X_{r,k} = 1$.

[5]For the two sources to be equally preferred when disregarding risk requires for equation (2.5) that $\beta_j = -\beta_{r,j}\left(\frac{2^{1-0}-1}{1-0}\right) = -\beta_{r,j}$.

Figure 2.2: Marginal utility of switching from source k to source j where source j is considered safer than source k and $\beta_{rj} < 0$.



choice of a new water source, is an empirical question that we seek to answer using the data described in the next section. To address this question we assign a subset h of sources with the risky intrinsic attribute the dummy variable $X_{r,h} = 2$. In line with the illustration above, we reject the null hypothesis that participants did not consider a particular type of risk in their choice of water source when $\beta_{r,h} \neq 0$.

We test three hypotheses using three different groupings of water supply sources: we assess the riskiness of each source individually as well as for a subset of sources that are subject to supply risk and another subset that is subject to technology risk. The first hypothesis tests whether the utility for any water supply source depends on risk. Empirically, we must set one source as the reference level. In our setting we test each water source relative to the omitted categorical variable 'new dam', which implies further development of Australia's conventional water supply source. Source j is considered less risky relative to new dam if $\beta_{r,j} < 0$ and riskier than new dam if $\beta_{r,j} > 0$. The second hypothesis is that supply risk is an important intrinsic attribute for the three weather-dependent sources: new dam, stormwater harvesting and interbasin transfer pipeline. To test if supply risk is an important intrinsic attribute we test the null against a one-sided alternative hypothesis that $\beta_{r,supply} > 0$ when $X_{r,supply} = 2$ for weather-dependent sources. The third hypothesis, following the literature on technology adoption and risk aversion (Liu, 2013; Qiu et al., 2014), is that new technology risk is an important intrinsic attribute of certain water sources. Recycled and stormwater harvesting are new technologies that are not widely used in Australia. All other sources have some well established and sizable capacity (Productivity Commission, 2011). Thus, we assign $X_{r,tech} = 2$ to recycled water and harvested stormwater and test whether water technology risk matters to households by defining the null hypothesis $\beta_{r,tech} = 0$ against the alternative that $\beta_{r,tech} > 0$. Our second and third hypotheses relate directly to the literature that identifies supply risk and the deployment of new technologies as the primary risks related to public water that concern the Australian public.[6] The objective of our study is to test whether these concerns affect householders' preferences for new sources of water supply in a fundamental way, and therefore, whether policy makers should focus their attention on these risks when discussing new water infrastructure investments in the public domain.

2.4 Survey Design and Data

2.4.1 Survey description

The discrete choice experiment (DCE) that elicits preferences for new water supply sources was part of a door-to-door survey on preferences for urban water management conducted in Melbourne and Sydney, Australia. In total, a random sample of 981 householders over the age of 18, who had owner-occupier status in 2013, were interviewed.[7] The sample was selected randomly by address, with eligible individuals who answered the door being invited to participate. Households were sampled from the council areas of Manningham and Moonee Valley (within greater Melbourne) and Fairfield and Warringah (greater Sydney). The councils were selected from 29 Cooperative Research Centre for Water

[6]For example, Dolnicar et al. (2014) show that broadly defined concerns about the safety and security/sustainability of water comprise 7 of the top 10 attributes of public water supplies. The list of desirable attributes, along with the percentage of respondents listing that attribute, can be found in Table 3 in Dolnicar et al. (2014).

[7]By only interviewing owner-occupiers we ensured that all participants in the survey also receive water bills, as this is not the case for some tenants.

Sensitive Cities (CRCWSC) partner communities, where local knowledge about water management is likely to be even higher than other communities in Australia.[8] The four councils were selected on the basis that they had similar rainfall patterns, income, age composition, level of home ownership and environmental preferences.[9] The survey was undertaken from March to October, 2013, ensuring results were not driven by seasonality.

At the door, interviewers introduced themselves and asked the householder to participate in a survey about local water management. The interviewer then confirmed the individual's eligibility, and proceeded with the survey on an iPad. Before commencing the survey, the software randomly assigned whether or not the participant would start by completing an incentivised risk experiment (approximately one in six of the total sample), with earnings for the task ranging from A$0.60 to A$23.10.

Next, respondents participated in a first DCE on the non-market benefits of local water management projects, described in more detail by Brent et al. (2014). The second DCE given to participants elicited water source preferences and is the focus of this paper. One third of respondents (those in the risk group and another group who was endowed with money at the start) had to pay for their decisions in the first DCE. No participants paid for their decisions in the second DCE, nor were they paid for participating in this task.[10] The survey ended with a set of demographic and water-relevant questions, after which participants without outstanding balances were paid in cash.

The survey was developed after a series of focus group meetings with researchers from different disciplines in the CRCWSC in which appropriate attributes and levels were discussed. A professional survey company was employed to conduct the survey, and the

[8]The CRCWSC is an Australian research organisation, which is funded by the federal government and has significant participation from partner communities and industry bodies.

[9]Data for the rainfall comparison were long-term mean and variance in daily, weekly and monthly precipitation, using daily rainfall data from January 1980 to February 2013, from the Australian Bureau of Meteorology. The other data were drawn from the Household, Income and Labour Dynamics in Australia (HILDA) survey, a government-funded panel study.

[10]While there are possibilities of order effects from the two DCEs we do not believe it will affect our central hypothesis about intrinsic risk attributes. Respondents in the risk task group, plus repondents in a second, similar sized group, were given an endowment and were required to pay for their responses for the first DCE. When comparing the responses of these treatment groups to the second DCE, we find no statistically significant differences between any of them. This is expected given that everyone faced the same hypothetical, non-incentivised choice sets for the second DCE. We also do not find any difference in responses between those who were given the risk task and those who were not.

interview team was carefully briefed by the authors with regards to the objective and details of the survey. The survey was then pre-tested in full length interviews with volunteer council employees, most of whom were not involved with water management in the council. A trained psychologist assisted the focus group interviews, conducted debriefing interviews with the participants and provided recommendations based on her assessment of the survey design (including wording, length, information content and cognitive demands). The revised survey was successfully tested in the field with a small sample of households before being rolled out.

Incentivised risk experiment

Before commencing the DCE a randomly selected subset of 167 respondents participated in a fully incentivised risk experiment involving choices over monetary lotteries, designed to allow risk attitudes to be estimated.[11] Experiments involving risk tasks are particularly useful for understanding how people make decisions involving risk (Charness et al., 2013) and have been utilised in areas such as understanding farmer adoption of new technology (Liu, 2013) and predicting health-related behaviours and preferences (Lusk and Coble, 2005; Anderson and Mellor, 2008). Furthermore, by fully incentivising the risk task we address concerns of hypothetical bias in the elicitation of risk attitudes (Holt and Laury, 2002; Lee and Hwang, 2016). Full instructions and explanatory examples shown to participants are given in Appendix A.3. The experiment is based on Holt and Laury (2002) and consists of ten questions, each of which asks the participant to choose between two binary lotteries.

The full set of questions are displayed in the first two columns of Table 2.2, which show the potential earnings and probabilities for each of the two lotteries. The third column of Table 2.2 shows the difference in expected value of lottery A and lottery B; the fourth gives the implied range for the coefficient of CRRA γ_i if the participant switches from lottery A to B at that question. The tenth question in Table 2.2 is a choice between

[11]Risk elicitation was a component of a randomized field experiment linked to the first DCE in the survey that is unrelated to the DCE over new water sources. As a result, the risk task was not rolled out over the entire sample.

Table 2.2: Risk preference task questions, difference in expected values and coefficient of CRRA.

Option A	Option B	$EV_A - EV_B$	CRRA if switch to B
10% of \$12.00, 90% of \$9.60	10% of \$23.10, 90% of \$0.60	\$6.99	$\gamma_i < -1.71$
20% of \$12.00, 80% of \$9.60	20% of \$23.10, 80% of \$0.60	\$4.98	$-1.71 < \gamma_i < -0.95$
30% of \$12.00, 70% of \$9.60	30% of \$23.10, 70% of \$0.60	\$2.97	$-0.95 < \gamma_i < -0.49$
40% of \$12.00, 60% of \$9.60	40% of \$23.10, 60% of \$0.60	\$0.96	$-0.49 < \gamma_i < -0.15$
50% of \$12.00, 50% of \$9.60	50% of \$23.10, 50% of \$0.60	−\$1.05	$-0.15 < \gamma_i < 0.15$
60% of \$12.00, 40% of \$9.60	60% of \$23.10, 40% of \$0.60	−\$3.06	$0.15 < \gamma_i < 0.41$
70% of \$12.00, 30% of \$9.60	70% of \$23.10, 30% of \$0.60	−\$5.07	$0.41 < \gamma_i < 0.68$
80% of \$12.00, 20% of \$9.60	80% of \$23.10, 20% of \$0.60	−\$7.08	$0.68 < \gamma_i < 0.97$
90% of \$12.00, 10% of \$9.60	90% of \$23.10, 10% of \$0.60	−\$9.09	$0.97 < \gamma_i < 1.37$
100% of \$12.00, 0% of \$9.60	100% of \$23.10, 0% of \$0.60	−\$11.10	$1.37 < \gamma_i$

receiving \$12.00 with certainty (option A) and \$23.10 with certainty (option B) and acts as a control question.[12] Before the task commenced, it was explained that one of the 10 questions would be randomly selected for payment. A random draw was used to determine which outcome of the selected option was paid to the participant.

To allow for more flexibility in the estimation of individuals' risk attitudes and to address concerns about order effects, we depart from Holt and Laury (2002) by presenting each question separately and in a random order rather than displaying the questions in a multiple price list format. This accommodates participants who display multiple switchpoints between lottery A and lottery B because they are indifferent between a number of lottery choices and thus their implied range of γ_i cannot be estimated as precisely as for respondents with single switchpoints (Andersen et al., 2006; Charness et al., 2013). For example, if a participant records lottery A for his first choice, lottery B for his second, lottery A for his third and lottery B thereafter, his estimated γ_i value lies between $-1.71 < \gamma_i < -0.15$.[13] Moreover, showing the lotteries to participants as a list (as in Holt and Laury, 2002) could lead to ordering effects that impact individuals' choices (Harrison et al., 2005; Dave et al., 2010). The randomisation of questions employed in this study may lead to noisier data, but is less likely to be biased. This increased noise will lead to an underestimation of the statistical significance of the impact of intrinsic risk

[12] A participant choosing option A for question 10 could imply that they do not wish to take money from the researcher or that they did not understand or engage with the task.

[13] This statement is made in accordance with the order of questions in Table 2.2, rather than referring to the particular random order in which the questions were displayed to the participant.

Figure 2.3: Example representation of the risk task to respondents.

on individuals' choices, thus the results presented in this paper are conservative in terms of measuring support for our hypotheses. To reduce the cognitive burden of respondents, all lottery payoffs and probabilities were presented using images as well as text (Dave et al., 2010), as shown in Figure 2.3.

Discrete choice experiment over water sources

Preferences for a new water supply source are elicited through a discrete choice experiment (Carson and Louviere, 2011). The task was introduced to participants as follows (full instructions are shown in Appendix A.4):

When water shortages become more frequent, investments to increase urban water supply need to be made. There are a number of options in terms of water source and technology that a city can invest in. These options differ with respect to the quality of water provided and therefore their allowed use, as well as the cost of water provision. It is possible to install a third water pipe to your house, so that your tap water will not be contaminated with potentially lower quality water from the new source. You would NOT have to pay for the installation of the third pipe.

You will now be asked to make a series of 10 choices regarding your preferred additional water source, its allowed uses and the resulting cost of water. Assume that this would be *the cost of your total water consumption per kiloliter in AUD. No other rates or charges would change.*

Before starting the DCE participants received a brief explanation about the different water sources and attributes. This explanation did not mention risk to ensure that the respondents' preferences over intrinsic risk attributes can be estimated without potential framing confounds. Throughout the choice task the participants could refer to the summary information sheet, which is reproduced in Figure A.2 of Appendix A.4. Each participant was then given a sequence of ten separate questions, presented in a graphical format. Figure 2.4 provides an example. Each question asked for the participant's preferred new water supply source out of six possibilities: desalination, recycled, new dam, groundwater, stormwater or pipeline (interbasin transfer).[14] As shown in Figure 2.4, the water supply source attributes vary in terms of allowed use and total cost per kiloliter on their water bill. Allowed use in the study has three levels – low risk outdoor use (non-potable outdoor, first two images, by descending order, in Figure 2.4); adding toilet, laundry and vegetable gardens (non-potable indoor, third image); and fully potable water (fourth image). Cost per kiloliter ranged from \$1.60 to \$3.20, in 20c increments. The lower cost levels were representative of water prices at the time of the survey while the higher levels are within realistic bounds.

The D-efficiency criterion was applied to construct four blocks of ten choice questions using the the software package *Ngene*. Each participant was randomly assigned to one of the four blocks, and they saw the questions from their given block in a random order. Overall, the questions were balanced so that each water source was assigned each level of allowed use and cost approximately the same number of times. New dam and desalination were only assigned the allowed use category of potable as this reflects the water quality most commonly supplied by these sources. All other new water supply sources could be assigned any of the three levels of allowed use.

The purpose of the survey is to determine community preferences over alternative

[14]While six alternatives may seem high in number for choice experiments, these are the six primary water sources available in Australia, and excluding any could introduce a bias into the respondents' choices. For example respondents may lump an omitted source together with one of the alternatives in the choice set.

Figure 2.4: Example of image shown to participants for a water supply source choice.

future water supply augmentations, conditional on a new water supply source being developed. Accordingly, this survey represents a forced choice, DCE as there is no "status quo" option for participants – for example "no new water source" (Hensher et al., 2005; Louviere et al., 2010; Carson and Louviere, 2011).[15] A status quo option such as "no new water source" brings with it implicit assumptions on the part of the participant about water supply reliability compared with building a new source. These implicit assumptions are not known to the researcher, making the interpretation of the results problematic. Respondents may associate a type of new water source with a known project, but the potential impact on their local amenities of a particular water supply source is a relevant consideration for them to be making. Thus, the method chosen represents the best method to elicit community preferences about options for centralised water supply augmentation

[15]Forced choice experiments are useful when considering situations such as preferences for the type of development in a place where a development is inevitable, and how residents value more conservation-friendly development (eg. Johnston et al., 2003; Duke et al., 2014). This study looks at a similar situation, asking participants to consider the inevitable situation in which not building a new water source is untenable.

(Hensher et al., 2005; Louviere et al., 2010; Carson and Louviere, 2011).

2.4.2 Descriptive statistics

The demographics, flood risk perception and flood insurance ownership of the full sample of 981 participants are recorded in Table 2.3.[16] The second to last column of Table 2.3 shows the same data for the subsample of 137 respondents for whom we have observed risk attitudes.[17] The rightmost column of Table 2.3 shows p-values comparing the distribution of each variable between the risk subsample and those in the full sample who are not in the risk subsample. The p-values are all well above 0.1, indicating the risk subsample is not statistically different from the full sample. Thus, conclusions drawn from the risk sub-sample are relevant for the whole sample. The overall choices made in the DCE are given in Figure A.1 in Appendix A.1.

Risk preference summary statistics

Table 2.4 shows the number of times each participant switched from the safe lottery A to the risky lottery B, using the order of questions in Table 2.2 as the order of lotteries.[18] Switching twice implies the person switched from lottery A to B at some point, then back to A, then to B again. As shown in Table 2.4, about half of the participants switched more than once. This is to be expected given participants saw the choices in a random order and thus were not biased towards having a single switch point, but rather could express indifference between some options by switching more than once (Andersen et al., 2006; Charness et al., 2013). Multiple switching is not uncommon even when using the original Holt and Laury (2002) multiple price list format, with Anderson and Mellor (2008) reporting 21% switching more than once from their large sample of the general population in the USA.

To utilise the estimated coefficients of CRRA in the modelling approach of this paper, we allocate the midpoint of the estimated range for γ_i (implied by their first switch,

[16]Flood risk perceptions and owning flood insurance are used to impute risk preferences as described in the Appendix.

[17]Thirty of the 167 who were given the risk task were excluded, as explained in Section 2.4.2.

[18]Answering lottery B for the first question of Table 2.2 is considered one switch.

Table 2.3: Summary statistics.

	Full sample (%)	Risk subsample (%)	p-value
Gender			0.294
Female	46.5	42.3	
Age			0.187
Refused	0.2	0	
18 to 24	4.0	5.8	
25 to 44	24.5	31	
45 to 64	41.7	46.7	
65+	29.7	24.8	
Education			0.500
Refused or other	4.0	1.5	
Year 10-12	27.3	24.8	
Certificate	15.3	16.8	
Associate	13.4	14.6	
Bachelor	23.8	21.2	
Graduate	16.3	21.2	
Income			0.608
Refused	4.1	3.0	
Don't know	2.6	0.7	
Low	23.2	22.2	
Middle	60.1	61.5	
High	10.0	12.6	
Flood risk perception			0.316
Refused	0.1	0	
Don't know	2.8	2.9	
1 in 2 years	7.2	4.4	
1 in 5 years	8.3	11.7	
1 in 10 years	8.4	9.5	
1 in 20 years	7.2	5.8	
Almost never	66.1	65.7	
Flood insurance			0.739
Refused	0.3	0	
Don't know	22.2	19.0	
Yes	38.1	38.7	
No	39.4	42.3	
Sample size	981	137	

Note: The p-values compare the risk sub-sample to the non-risk participants in the full sample, by variable, using Pearson's χ^2 test to compare the distribution of categorical variables, except for age. As we have data on exact age, a two-sided non-parametric Kolmogorov-Smirnov test for differences in distributions is used given the χ^2 test performs poorly when there are many categories relative to sample size.

Table 2.4: Number of switches between lotteries A and B.

Number of switches from A to B	**% of participants**
1	49.6
2	33.6
3	13.1
4	3.6
Sample size	137

and last switch if they switch more than once as explained in Section 2.4.1) to each participant (see Andersen et al., 2006; Liu, 2013, and others who use this method). We use a conservative approach to deal with issues of unboundedness and use a γ_i parameter value of -1.71 for people who selected option B in the first question and 1.37 for people who switched from option A to option B for the last question.[19]

Of the 167 participants who completed the risk task, we exclude 30 who chose option A for question 10 since they may not have understood the risk task.[20] The 137 participants for whom risk attitude is observed are a random subsample of the full 981 participants, as already shown in Table 2.3. The median, mean and standard deviation of the observed coefficient of CRRA are 0.21, 0.10 and 0.88 respectively. This shows that the majority of participants are risk averse, but with considerable heterogeneity, as found in similar field experiments (Anderson and Mellor, 2008; Harrison et al., 2007; Dave et al., 2010). However, there are a couple of differences in our results compared with other studies that are worth noting.

First, the mean coefficient of CRRA we find is not statistically different from zero. However, the median is higher than the mean, and the median is within the coefficient of CRRA range implied by switching at choice six in Table 2.2, which is firmly in the risk averse range. The mean is drawn down by the fact that the lowest possible value of the CRRA coefficient is -1.71, while the highest is 1.37. There are some participants at these extremes (9% and 12% respectively, and with 12% below -1.37). In fact, there is less

[19] An alternative would be to assume a lower and upper bound based on the most extreme values found in the literature. Experimentation with this alternative approach did not yield material differences to the overall results of this study. Also, the majority of Danes in a similar field study were found to exhibit a CRRA parameter within the range of -1.71 and 1.37 (Harrison et al., 2007).

[20] The relatively large number of respondents included is likely due to our deviation from the multiple price list format, which we do in order to reduce single-switching bias.

bunching around the middle of the distribution compared with other studies. This second difference is likely due to the random ordering of the questions, and is an indicator that there may be a bias toward switching at the middle created when showing the options in a list format. Hence, this finding supports our decision to randomly order the questions. Holt and Laury (2002) and Dave et al. (2010) find 75% and 69% of switch points within questions 4 to 6, implying a CRRA coefficient of $-0.49 < \gamma_i < 0.41$, whereas we find less of a bias towards the middle, with 35% of participants falling within that range.

2.5 Empirical Specification

This paper employs the mixed logit to estimate the utility function given by equations (2.3) and (2.4). An advantage of the mixed logit is that it allows for preference heterogeneity among participants, by incorporating both fixed and random coefficients.

To simplify notation we group all coefficients into a single vector $\boldsymbol{\beta}$, and all variables for source j at time t into a single $\boldsymbol{X}_{jt}$. U_{ijt} can be modeled probabilistically, as it is a latent variable that determines each individual's choice of water supply source, j. Thus, assuming each individual has a unique $\boldsymbol{\beta}_i$

$$\begin{aligned} \Pr(Y_{it} = j) &= \Pr(U_{ijt} > U_{ikt}) \quad (\forall j \neq k) && (2.6a) \\ &= \Pr(\boldsymbol{\beta}_i \boldsymbol{X}_{jt} + \epsilon_{ijt} > \boldsymbol{\beta}_i \boldsymbol{X}_{kt} + \epsilon_{ikt}) \quad (\forall j \neq k) && (2.6b) \\ &= \Pr(\epsilon_{ikt} - \epsilon_{ijt} < \boldsymbol{\beta}_i \boldsymbol{X}_{jt} - \boldsymbol{\beta}_i \boldsymbol{X}_{kt}) \quad (\forall j \neq k). && (2.6c) \end{aligned}$$

As the objective is to compare models that explicitly allow for water source specific risks with those that do not and for which the error terms would be correlated, we reject the independent and identically distributed (IID) assumption and specify a mixed logit functional form for equation (2.6c). The mixed logit model allows for individual heterogeneity in $\boldsymbol{\beta}$ in the following way:

$$\Pr(Y_t = j) = \int \frac{\exp(\boldsymbol{\beta}\boldsymbol{X}_{jt})}{\sum_{k \in J} \exp(\boldsymbol{\beta}\boldsymbol{X}_{kt})} f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta}. \tag{2.7}$$

Here, $\boldsymbol{\theta}$ is a vector of distributional parameters, such as the mean and variance, estimated using numerical simulation of maximum likelihood. Estimating the model requires the specification of the distribution of each element of $\boldsymbol{\beta}$, and whether or not they are independently distributed, or correlated. Commonly normal, lognormal or triangular distributions are used. By allowing random distribution of $\boldsymbol{\beta}$, the mixed logit can approximate any random utility model (Hensher and Greene, 2003; Train, 2009).

2.6 Results

The base model in the first column of Table 2.5, is based on equation (2.3) and is the mixed logit estimation of the explicit, extrinsic attributes presented in the DCE. It is estimated on a subsample of 860 people using maximum simulated likelihood with 400 Halton draws; this number of draws is used to ensure stability of estimates for this dataset and model specification (Hensher and Greene, 2003; Train, 2009).[21] The first two coefficients in descending order are fixed coefficients for allowed use – non-potable outdoor and non-potable indoor, relative to potable quality. The results confirm findings in other studies that people dislike non-potable indoor water. Chen et al. (2013) accredit this aversion to concerns over smell and colour of this type of water, given it is used for toilets and laundering. While other specifications were tested, the goodness of fit measures of AIC and BIC indicate that the quality coefficients should be fixed.

The next set of variables in column (1) of Table 2.5 are the means of the random ASC coefficients for water source, relative to new dam. The coefficients on these variables are in line with the overall choices (see Figure A.1 in Appendix A.1): they are all negative as new dam is the most popular option. Desalination, with the largest mean ASC, is the next most preferred source and the groundwater ASC is the smallest indicating that it is the least popular source. All water source coefficients are assumed to be normally

[21]The subsample of 860 is used so that it the same subsample as all models in Table 2.5, which arises as a result of the imputation process. This is explained in detail in Appendix A.2.

Table 2.5: Mixed logit regression results.

	Base	All with risk	Supply Risk	Technology Risk
	(1)	(2)	(3)	(4)
Fixed Coefficients & Means				
Fixed Coefficients				
Non-potable outdoor	0.0265	0.0259	0.0259	0.0259
	(0.0470)	(0.0511)	(0.0496)	(0.0481)
Non-potable indoor	-0.1452^{***}	-0.1471^{***}	-0.1471^{***}	-0.1455^{***}
	(0.0514)	(0.0504)	(0.0531)	(0.0498)
$\beta_{r,desalination}$		-1.2484^{***}		
		(0.4682)		
$\beta_{r,recycled}$		-0.7858		
		(0.6179)		
$\beta_{r,groundwater}$		-0.2815		
		(0.4957)		
$\beta_{r,stormwater}$		-0.3155		
		(0.5083)		
$\beta_{r,pipeline}$		-0.1122		
		(0.4029)		
$\beta_{r,supply}$			0.7115^{*}	
			(0.3847)	
$\beta_{r,tech}$				-0.3891
				(0.4581)
Random Coefficients				
Desalination	-0.7724^{***}	0.4811	-0.0546	-0.7746^{***}
	(0.0879)	(0.4661)	(0.4014)	(0.1021)
Recycled	-1.6845^{***}	-0.8863	-0.9622^{**}	-1.2903^{***}
	(0.1109)	(0.6392)	(0.3995)	(0.4823)
Groundwater	-2.5589^{***}	-2.2713^{***}	-1.8375^{***}	-2.5616^{***}
	(0.1207)	(0.5202)	(0.4047)	(0.1331)
Stormwater	-0.9977^{***}	-0.6797	-0.9998^{***}	-0.6053
	(0.0788)	(0.5250)	(0.0845)	(0.4747)
Pipeline	-2.2565^{***}	-2.1380^{***}	-2.2534^{***}	-2.2554^{***}
	(0.0980)	(0.4220)	(0.0992)	(0.1074)
Cost	-0.1118^{***}	-0.1073	-0.1086	-0.1138
	(0.0425)	(0.0884)	(0.0927)	(0.0912)
Standard Deviation or Spread				
Standard Deviation				
Desalination	2.1183^{***}	2.0891^{***}	2.0923^{***}	2.1244^{***}
	(0.0961)	(0.1020)	(0.1025)	(0.1068)
Recycled	2.2761^{***}	2.2566^{***}	2.2593^{***}	2.2716^{***}
	(0.1083)	(0.1192)	(0.1205)	(0.1223)
Groundwater	1.6403^{***}	1.6369^{***}	1.6346^{***}	1.6458^{***}
	(0.1013)	(0.1381)	(0.1276)	(0.1334)
Stormwater	1.6482^{***}	1.6492^{***}	1.6516^{***}	1.6520^{***}
	(0.0729)	(0.0926)	(0.0940)	(0.0877)
Pipeline	1.3142^{***}	1.3089^{***}	1.3092^{***}	1.3118^{***}
	(0.0910)	(0.1409)	(0.1321)	(0.1306)
Spread				
Cost	0.2639^{***}	0.2522^{***}	0.2549^{**}	0.2701^{**}
	(0.0981)	(0.0946)	(0.1005)	(0.1058)
AIC	23795.0	23790.4	23787.7	23794.2
BIC	23893.8	23924.5	23893.6	23900.1
Observations	8600	8600	8600	8600
Individuals	860	860	860	860

Note: Standard errors clustered at the respondent level are in parentheses. CRRA data is imputed for 723 individuals for models (2) to (4), and thus the standard errors are bootstrapped for these models. The coefficient for cost follows a triangular distribution. All other random coefficients are normally distributed. Allowed use variables are relative to potable, water source variables are relative to new dam. All models are estimated using 400 Halton draws. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

distributed.

The final random coefficient is cost. The mean is negative and statistically significant, as expected. Using a symmetric triangular distribution, we find that sensitivity to cost is low but within a reasonable range. Sensitivity to cost is often low when using realistic values for water given these costs are low compared with a total household budget (Olmstead, 2010). We use an unbounded triangular distribution that allows more flexibility to account for this fact.

The next section of the table shows the standard deviation or spread of the random coefficients. The estimated standard deviation for the new sources of water coefficients are large and significant. Thus, preferences for new water source are highly heterogeneous. The spread of the cost coefficient is also significant, indicating a range of cost sensitivities among respondents.

2.6.1 Incorporating preferences for intrinsically risky attributes

The three hypotheses regarding source specific risk, supply risk and technology risk are tested subsequently in models (2)-(4). In order to utilise as many individuals in the sample as possible, we use imputed risk preference data. The imputation involves regressing demographic variables and indicators of attitudes to risk on the observed CRRA parameter. These are jointly significant at the 1 per cent level. The fitted values from this approach are used to impute the risk attitudes of the 723 people who did not participate in the risk task and for whom we have observations on all the relevant variables for the imputation. The mean CRRA parameter value and standard deviation of the full dataset of 860 respondents with either observed or imputed CRRA parameter values, is 0.08 and 0.58 respectively. This compares favourably to the mean and standard deviation of 0.10 and 0.88 for the observed sample. The results of the imputation and further details are presented in Appendix A.2; these details include Table A.2, which estimates the models in 2.5 using just individuals with observed risk preferences. The results are very similar overall, but yield slightly lower levels of statistical significance for the coefficients due to the smaller sample size.

Bootstrapping of standard errors is undertaken in all models (2) to (4) of Table 2.5 in order to account for the uncertainty from the imputation stage. We use the Shao and Sitter (1996) method for bootstrapping, as it is robust to imputation method. It requires the full imputation procedure to be completed for each bootstrap replication. As a slight departure from Shao and Sitter (1996), we split the sample into those 137 individuals with observed risk attitudes and those 723 individuals with unobserved risk attitudes and we sample each separately, with replacement. This split bootstrap sampling is done to reflect the original survey design. Because of the random allocation of the risk task among the survey participants, this split bootstrap sampling process does not impact the validity of the estimated standard errors.

In model (2) of Table 2.5 we test the first hypothesis of whether the utility for any water supply source depends on risk. Thus, we estimate the model from equation (2.4) with a vector of risk dummies $\boldsymbol{X}_r$, such that each source j except new dam has a unique $\beta_{r,j}$. We conduct a two-sided test on each $\beta_{r,j}$; further we note that if $\beta_{r,j} < 0$ the source is considered safe relative to new dam, and risky if $\beta_{r,j} > 0$.

We find that only $\beta_{r,desalination}$ is individually, statistically significant and different from 0 (at the 1% level). Thus, only the intrinsic risk profile of desalination is found to be significantly different from that of a new dam. Specifically, the negative sign on $\beta_{r,desalination}$ indicates that augmenting the water supply with desalinated water is considered less risky than sourcing additional water from a new dam. This result is intuitive in light of the frequent water shortages that are imposed in Australia as a result of the reservoirs' vulnerability to droughts. Desalination, on the other hand, is seen as the most robust, drought-resistant supply source. Ranking all sources by the size of their $\beta_{r,j}$ coefficient and ignoring statistical significance for the moment reveals that desalination is perceived to be the least risky source, followed by recycled, stormwater, groundwater, pipeline and finally new dam.

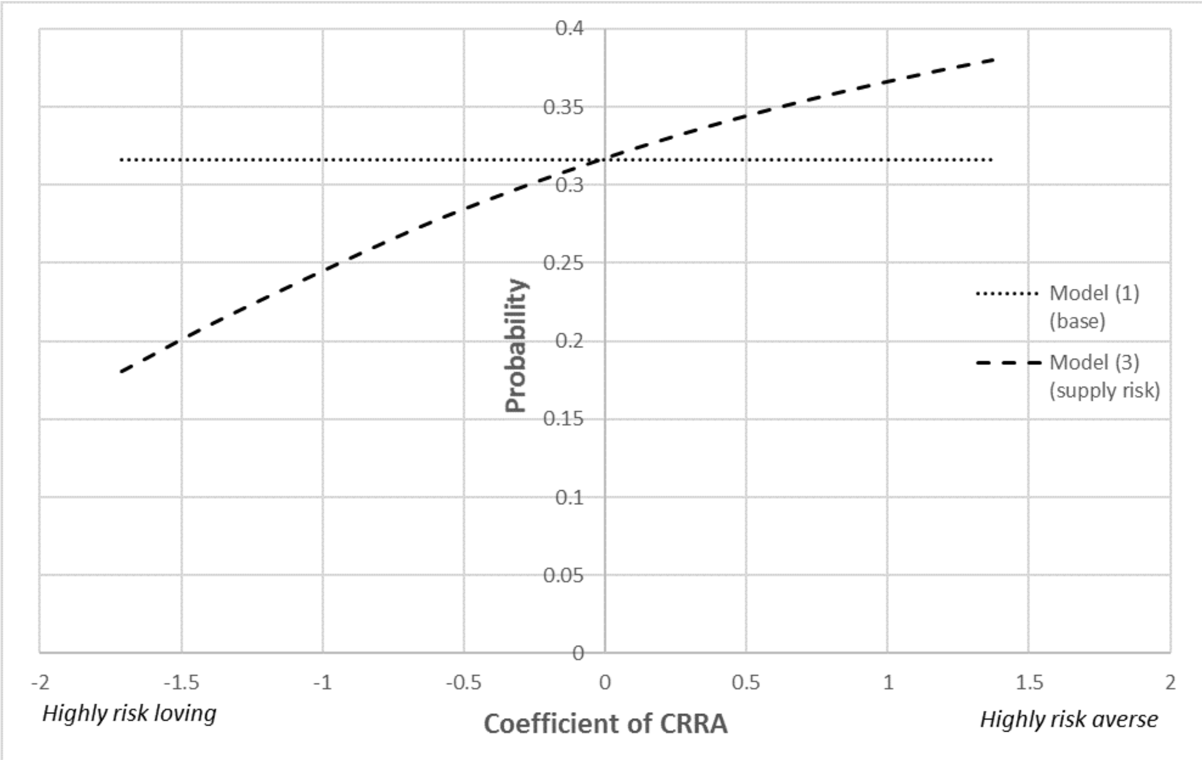
Following from equation (2.5), both the ASCs and the $\beta_{r,j}$ coefficients must be taken into account when comparing preferences for sources in model (2), and for any model with intrinsic risk. In the base model, the only relevant coefficients for comparing preferences, *ceteris paribus*, are the ASCs. As an example, the difference in model (2) relative to model (1) can be observed for the ASC for desalination. This coefficient goes from negative and statistically significant in in the base model, to positive and insignificant in model (2). However, taking into account $\beta_{r,desalination}$ and risk preferences, overall new dam is still preferred to desalination at the mean in model (2) as in model (1). The difference is that the results for model (2) can be used to determine how risk aversion affects the preferences for desalination relative to new dam.

Supply risk preferences

In model (3) of Table 2.5 we test the second hypothesis that supply risk is an important intrinsic attribute. We assign the three weather-dependent sources (new dams, stormwater harvesting and interbasin-transfer pipeline) the risk variable $X_{r,supply} = 2$ and formally test the null that $\beta_{r,supply} = 0$ against the alternative that $\beta_{r,supply} > 0$. Using a one-sided test, we reject the null hypothesis in favour of the alternative hypothesis at the 5% level. Furthermore, the model fit improves over the base model (1) and over model (2) using both AIC and BIC criteria. Combining this result with the estimated β_r coefficients in model (2) that ranked new dam, pipeline and stormwater respectively as first, second and fourth riskiest sources, we conclude that supply risk is an important driver of preferences for weather-dependent sources. While the results from model (2) suggest that it is the supply risk of new dam relative to desalination that is a major driver behind the supply risk coefficient, it is important to model supply risk as a single joint coefficient to test whether supply risk is an overall driver of preferences. Similar to the interpretation of individual source risk, the marginal utility from supply risk must be taken into account in addition to the ASCs when comparing preferences for sources in model (3).

Accounting for supply risk has important consequences for the probability with which a specific source is preferred over another. For example, model (3) predicts a risk loving individual is 34% more likely to choose new dam compared with a risk averse individual. This result is reversed for desalination, where a risk loving individual is 52% less likely to choose it compared with a risk averse individual. Figure 2.5 shows the probability of

Figure 2.5: Probability of choosing desalination over new dam by level of risk aversion for the base model (1) and when accounting for supply risk (model 3).



choosing desalination over new dam as predicted by models (1) and (3). As can be seen, the probability predicted by the base model in Table 2.5 does not vary by risk preferences. In contrast, in model (3) the probability of choosing desalination over new dam more than doubles from a highly risk loving to a highly risk averse individual.[22]

New technology risk preferences

The third hypothesis concerning the importance of technology risk in driving preferences is tested in model (4). Here we assign $X_{r,tech} = 2$ to the new and unfamiliar technologies (recycled and stormwater) and formally test the null hypothesis that $\beta_{r,tech} = 0$ against the alternative that $\beta_{r,tech} > 0$. The estimated coefficient of $\beta_{r,tech}$ for new technology risk is negative and statistically insignificant, thus indicating that the null hypothesis cannot be rejected. We therefore conclude that new technology risk is not an important driver of preferences over additional sources of municipal water. Furthermore, using the AIC and

[22] All the values in this paragraph and Figure 2.5 are calculated *ceteris paribus*, assuming ASCs at their means and all sources are of high quality and cost \$2.40 per kiloliter. Risk preferences used are at the extreme CRRA values of -1.71 and 1.37.

BIC criteria, the model incorporating technology risk does not fit the data as well as the model with supply risk.

2.7 Conclusion

Preferences drive choices, and incorporating parameters such as risk attitudes into choice modelling produces a more comprehensive picture of preferences in a given setting. In this paper we demonstrate how data on risk preferences can disentangle the importance of specific intrinsic attributes in driving preferences for a particular type of good.

When using DCEs to elicit community preferences for non-market goods, risk often plays a central role in determining the optimal allocation of resources. Some recent studies that allow risk to vary explicitly find that risk matters for preferences. However, what truly drives decisions is risk perceptions, which may or may not be related to the defined risk levels in a DCE. Moreover, if existing perceptions about an attribute are well-established the attribute cannot plausibly be varied across alternatives. Additionally, there is a limit to how many attributes can be included in a DCE experiment before cognitive limits are reached. We demonstrate that measuring attitudes towards intrinsic attributes can help identify which, and to what extent, intrinsic attributes drive preferences. This approach can be generalised to account for other experimentally-elicited preferences such conditional cooperation and trust.

We utilise a fully incentivised risk experiment to rigorously elicit risk attitudes of respondents. We leverage this information on risk attitudes to model the intrinsic risk perceptions and preferences over new water supply sources in a setting where the public knowledge about water source risk is high. Indeed, the respondents in our sample frequently experience water restrictions imposed by water shortages and are subjected to many high profile public debates regarding water supply augmentation options. By extending a basic random utility model to incorporate observed and imputed risk attitudes, we are able to test whether water supply risk and new technology risk are important to participants. We find no evidence that technology risk is an important consideration when choosing alternative sources of municipal water supply. In contrast our results suggest that water supply risk is an important driver of preferences and that including this type of intrinsic risk improves model fit. These findings are important for water managers who want to utilise green infrastructure for water management but are concerned about the public perception of alternative supply sources.

Chapter 3

Intrinsic motivation, health outcomes and the crowding out effect of temporary extrinsic incentives: A lab-in-the-field experiment

3.1 Introduction

Extrinsic incentives, including monetary incentives, have been used or are being considered for use to change behaviour in a number of areas of public policy. Examples include for health, environment and education (Gneezy et al., 2011). Within the health domain, there is international policy interest in paying people to increase or change their health related behaviours, including payment to quit smoking, exercise and attend disease screenings (Promberger and Marteau, 2013). However, evidence on the potential for extrinsic incentives, particularly monetary incentives, to reduce rather than increase effort has been gathering since Deci's seminal article in the field of psychology (Deci, 1971; Gneezy et al., 2011). Intrinsic motivation to do an activity can take many forms, such as the enjoyment of a particular activity, the desire to engage in productive and meaningful work, the benefits to one's self image from undertaking an activity or prosocial motivation

(Promberger and Marteau, 2013). When intrinsic motivation is crowded out, one or both of the following occurs: effort is reduced after the application of incentives; or effort is reduced below pre-incentive levels after the removal of temporary incentives. A number of laboratory experiments have investigated responses to extrinsic incentives within the former context, comparing effort level between subjects in response to extrinsic incentives (eg. DellaVigna and Pope, 2017; Heyman and Ariely, 2004; Pokorny, 2008). Less attention has been paid to temporary extrinsic incentives, which could be used to encourage the formation of habits such as healthy behaviours.[1]

The aims of this paper are to better understand intrinsic motivation and its importance for health outcomes, and the impact of type and size of temporary extrinsic incentives on intrinsic motivation. We undertake a lab-in-the-field study on a heterogeneous adult population. This approach allows us to carefully control the setting, while being able to measure the importance of heterogeneity and improve external validity of the findings compared with a more standard, homogeneous undergraduate student subject pool.[2] Our application to health outcomes also benefits from a heterogeneous population, and demonstrates not just the external validity, but also the policy relevance of our experiment.

We employ a rich within and between subject design to test the impact of a range of incentives that vary by size and type, both during the application of the incentives and after the incentives are removed. We give subjects four time-limited rounds of a real effort task (based on Erkal et al., 2011). Existing literature suggests that real effort tasks provide subjects with utility and are also designed to give a relatively fine measure of effort on the intensive margin (Brüggen and Strobel, 2007; Gill and Prowse, 2012). In the first round, no incentives for effort are provided. In the second round, all but the control group are given an unexpected extrinsic piece rate incentive, which varies between treatment groups by size and type. In the third round the incentives are unexpectedly removed. The fourth round is also without incentives, and given to subjects after a break

[1] See Charness and Gneezy (2009) and Royer et al. (2015) for field experiments on using extrinsic incentives to form exercise habits.

[2] While the study is undertaken in a standard university laboratory, our non-standard subject pool qualifies the study as a lab-in-the-field, or an artefactual field experiment (Harrison and List, 2004; Viceisza, 2016).

to test persistence of effects in the third round.

We find a low power monetary incentive is the only incentive effective at increasing effort in round 2, whereas the high power monetary incentive only serves to crowd out effort in the subsequent rounds, particularly among those with the highest levels of original intrinsic motivation. This result fits with Pokorny's (2008) rule of "pay – but do not pay too much". However, the high power incentive is effective at increasing effort among those with lower initial intrinsic motivation and does not significantly crowd out intrinsic motivation within this group. Our findings for this group are more aligned with Gneezy and Rustichini's (2000b) "pay enough or don't pay at all". Finally, intrinsic motivation is found to partially explain waist-to-height ratio, which is a powerful indicator of general health (Ashwell and Hsieh, 2005).

Our major contributions to the literature stem from our baseline measure of intrinsic motivation, given by effort level in round 1. To the best of our knowledge, we are the first to test whether a laboratory measure of intrinsic motivation predicts health outcomes. The second major contribution of our study design relates to how we measure the impact of incentives. Generally, laboratory studies have compared incentives between subjects for a single round (eg. DellaVigna and Pope, 2017; Heyman and Ariely, 2004; Pokorny, 2008; Takahashi et al., 2016). Conversely, Deci (1971) uses a three round design, allowing for within subject effects. He compares the differences-in-differences between rounds 1 and 3 for a control group with no incentives, and a treatment group with an incentive in round 2.[3] We combine elements of both designs for our within and between subject approach, testing a range of incentives in round 2, while accounting for initial intrinsic motivation. We then test for crowding out by each incentive type when incentives are removed in round 3, allowing us to better understand the balance between the positive effects of the incentive and the negative effects of crowding out. Our design also provides us with the means to analyse heterogeneity in response to incentives by level of initial motivation. Finally, we contribute to the literature by adding a fourth unincentivised round to test for persistence of effects from round 3.

[3] See Ma et al. (2014) for a more recent study in the field of neuroscience with this design.

This paper is organised as follows. The second section below covers the background literature in more detail. In Section 3.3 we outline the theoretical framework, experimental procedures and design, the hypotheses we test and our analytical approach. The results are presented in Section 3.4, while Sections 3.5 and 3.6 provide a discussion and conclusion.

3.2 Background

The purpose of this section is to outline in more detail the literature that this paper fits within and our contribution. We begin by defining intrinsic motivation and extrinsic incentives, followed by a discussion of applications in the literature and where this paper fits. Next, we summarise the experimental studies undertaken on this topic, finishing with our key contributions.

In this paper, we use Bénabou and Tirole's (2003, p. 490) definition of intrinsic motivation, "the individual's desire to perform the task for its own sake". This definition is useful as it notes the three important features of our context. First, the individual's desire to undertake the task. Following Bénabou and Tirole's (2003) theory, we include the individual's belief in their own abilities within their desire to undertake the task. Second, the definition includes the nature of the task itself, such as how enjoyable or useful it is to the individual. Third, undertaking the task for its own sake is taken to mean the level of effort put into a task in absence of an extrinsic incentive. Understanding these three elements of the definition allows us to unpack the literature, as well as nicely framing our research.[4]

To define extrinsic incentive, we follow Cerasoli et al. (2014, p. 981): "anything provided by an external agent contingent on performance of particular standards of behaviour(s)." Hence, we use piece rate incentives to test extrinsic incentives. Another important distinction to make is type of incentive. We use both monetary and non-monetary

[4]Promberger and Marteau (2013) summarise the psychological literature's definition of intrinsic motivation as "Doing the task exclusively for its own sake." While the differences between this definition and Bénabou and Tirole's (2003) definition might be subtle, they are important. The major difference is that the psychology definition does not note the role of the individual in intrinsic motivation. On the other hand, while the economic literature also discusses intrinsic motivation, there is generally no clear definition given. Rather, there is a focus on cases where extrinsic incentives have the opposite effect to what neoclassical theory predicts, especially for prosocial activities (Promberger and Marteau, 2013).

incentives. Monetary incentives can have a crowding effect relative to non-monetary incentives of the same value (Heyman and Ariely, 2004). We test a charity payment, expressed as trees planted, as a non-monetary incentive, that is given in a piece rate fashion. Finally, we assume that a valid measure of intrinsic motivation is the observed effort level for an activity that is without an explicit extrinsic incentive.

Broadly, the nature of the tasks considered by the literature fit into two main categories – tasks with primarily private benefits, and tasks with primarily social benefits. Contexts with private benefits include enjoyable puzzles (Deci, 1971), health (Charness and Gneezy, 2009; Royer et al., 2015), education (Bettinger, 2012) and principal-agent contracts (Gneezy and Rustichini, 2000a). Contexts with social benefits include the environment (Brent et al., 2017; Kerr et al., 2012) and volunteering or incurring a private cost for the public good (Lacetera et al., 2014; Frey and Oberholzer-Gee, 1997). Some authors, such as Grant (2008) and Kamenica (2012), distinguish between the former and the latter by referring to the former as intrinsic motivation, and the latter as prosocial behaviours. Our research context of interest is health, hence our experiment has been designed for research within the private benefit context. The effort task we utilise is essentially a puzzle. A range of useful research into intrinsic motivation and extrinsic incentives has been undertaken within the laboratory context (Promberger and Marteau, 2013), but to our knowledge we are the first to directly test whether intrinsic motivation measured in the laboratory can predict health outcomes.

The observation that the economics literature on intrinsic motivation is focussed on prosocial or moral activities is as applicable to the theoretical literature as it is to the empirical literature. Widely cited theory papers for which this claim is true include Brekke et al. (2003), Bénabou and Tirole (2006), Bénabou and Tirole (2011) and Bowles and Polanía-Reyes (2012). A notable exception is Bénabou and Tirole (2003), for which the focus is on a principal-agent relationship, where the principal could range from being an employer to a teacher or health practitioner, and thus the relevant agent being an employee, student or patient. Therefore, beyond just using their definition of intrinsic motivation, we are also guided by Bénabou and Tirole (2003) to develop our experiment and interpret our results.

In terms of the psychology literature, the emphasis of studies on intrinsic motivation has been on motivation to do enjoyable activities – that is, within the realm of activities with private benefits. Effort in these enjoyable activities are crowded out by tangible incentives, such as money or tasty food (Deci et al., 1999; Promberger and Marteau, 2013). However, Cameron et al. (2001) argue that intrinsic motivation should be approached with a more broad definition than just enjoyable activities. When applying this broad definition, the literature shows that crowding out is not pervasive, but applicable only to high interest tasks with tangible rewards that are at least loosely performance-based.[5]

Consistent with Bénabou and Tirole's (2003) definition, we take a broad view of intrinsic motivation. Intrinsic motivation includes both any enjoyment gained from the activity, as well as reasons for putting in effort related to the nature of the individual, as noted at the start of this section. An individual may choose to put in effort as they are optimistic about the nature of the tasks and its benefits to themselves. Other reasons include self-competition and self-improvement (Khalil, 1997). Thus, we seek to deepen the understanding of intrinsic motivation within the empirical literature by considering whether there might be a general underlying level of intrinsic motivation, related to an individual's tendency towards self-belief and applying themselves to the task at hand. This is one of the motivations behind testing whether our measure of intrinsic motivation predicts real world health outcomes.

A range of experiments have been undertaken to test how individuals respond to extrinsic incentives for tasks where intrinsic motivation is a strong driver of effort (Promberger and Marteau, 2013). These experiments follow in the footsteps of Deci's (1971) classic result that monetary incentives crowd out intrinsic motivation, whereas verbal affirmations crowd in intrinsic motivation. However, in the monetary space, Deci (1971) only tested a monetary incentive against no incentive. Gneezy and Rustichini (2000b) test different levels of monetary incentives, and develop the rule of thumb to "pay enough or don't pay at all". Their low power monetary incentive is less effective than no incentive, but

[5]For the response to Cameron et al. (2001), see Deci et al. (2001).

their high power incentive increases effort the most. They argue the crowding out effect of money can be combatted with a high enough pay rate. Heyman and Ariely (2004) essentially reproduce the same result, except their high power incentive has roughly the same effect as no incentive.

On the other hand, Pokorny (2008) finds the low power incentives to be the most effective at raising effort, hence her suggestion to "pay – but do not pay too much". She explains her results with reference dependent preferences, arguing the student subjects are trying to earn a given amount based on what they expect from an experimental session. Another potential explanation for such a finding, perhaps when the task is difficult enough, is a choking effect that occurs when stakes are high (Kamenica, 2012). Takahashi et al. (2016) reproduce Pokorny (2008) using one task, and the standard economic prediction of a monotonic response to incentives using another task. They argue that this difference in their results is due to the different nature of the tasks. They contend that their first task, a computerised ball dragging task (similar to Heyman and Ariely, 2004), is "boring" and their second task, a puzzle, is more interesting. For a minimum threshold incentive with a piece rate, Kajackaite and Werner (2015) find subject effort declines after meeting a minimum threshold relative to those without a threshold. DellaVigna and Pope (2017) test a range of levels of monetary incentives and find a standard monotonic response. Their subjects show intrinsic motivation by putting in significant effort for the task in the no incentive treatment. The task is a simple real effort task of pressing two buttons as quickly as possible.

Studies have also looked at non-monetary incentives. DellaVigna and Pope (2017) find comparable effect sizes between monetary and non-monetary incentives, though overall monetary incentives are most effective. Heyman and Ariely (2004) find candy can be more powerful than a low power monetary incentive. Imas (2014) find low power charity incentives are more effective than low power monetary incentives of the same value, but the difference disappears for high power incentives.[6]

[6] Some studies, such as Tonin and Vlassopoulos (2014), look at a combination of private and charity incentives. We do not cover these studies in any detail as we compare a pure charity incentive to pure monetary incentives.

Other than Deci (1971), all these studies compare effort in response to treatments between subjects, for one round of the task. Hence, we build on the strengths of Deci's (1971) three round design of no incentive, incentive, no incentive, while also testing a range of incentive types and sizes between subjects. Following Deci (1971), we include a control group with no incentives for any round. This approach allows us to more deeply understand how extrinsic incentives affect intrinsic motivation, by including initial level of intrinsic motivation for the task, and allowing us to measure crowding out that has occurred once the incentive has been removed. We can thus better understand the balance between the positive impact of the incentive and the negative impact of crowding out. We use a real effort task to make our results comparable with similar studies. We test a low power and a high power monetary incentive, a high power monetary incentive with a threshold, and a charity incentive, expressed as a piece rate that plants native trees locally, but with the same monetary value as the high power incentive.[7] Thus, we can test whether going big or small with monetary incentives is shown to be a better strategy, and how alternative incentives perform, once we take into account initial intrinsic motivation. Finally, we use a non-standard subject pool to allow for greater heterogeneity of intrinsic motivation and responses to the incentives.

3.3 Method

3.3.1 Theoretical framework

We apply the theoretical model of Bénabou and Tirole (2003) to our research aims, as it forms a useful basis to guide our experimental design, hypotheses and the interpretation of our results. We discuss the basic and relevant features of their model here, and return to their notation to develop our hypotheses. We specifically consider their model in a multiple-round context where there are repeated interactions between the same principal and agent, which is not a focus of the original paper. Also additional to the original

[7] We chose the charity incentive for the non-monetary incentive as we are interested whether a charity donation can be an effective motivator in a private benefit context. Charitable donations are sometimes used to encourage exercise, for example fundraising "fun runs".

paper, we consider whether intrinsic motivation is correlated across domains.

The model is useful for our context as it is consistent with much of the literature on intrinsic motivation within contexts with primarily private benefits. Furthermore, it is applicable to our experimental context where there is a principal (the experimenter, "him"), and each subject represents an individual agent ("her"). The principal interacts with each agent individually for multiple rounds; there is no interaction between agents. Furthermore, the theoretical framework is highly applicable to related contexts, such as a government incentive programme to encourage exercise, where the government is the principal and citizens are the agents.

Bénabou and Tirole's (2003) model includes a principal's payoff function, $U_P(\beta, e, p)$, and an agent's payoff function, $U_A(\beta, e, p)$. The term e represents the level of effort the agent exerts and p is the treatment policy that the principal chooses in order to reward effort (the extrinsic incentive). The model centres around the term β, which is some information that the principal holds about the task or the agent's capacity to perform it, about which the agent is uncertain.[8]

The component of β that is the nature of the task could include the private benefits the agent will derive from it. The principal may have more information about how enjoyable the task is if the agent is unfamiliar with the task, and short and long term benefits that will accrue to the agent. In terms of the experiment in this paper, the agent (subject) is unfamiliar with the task, how enjoyable it could be, and whether there are any other benefits to undertaking it such as a mental workout. In the case of an exercise programme, an agent may not have exercised much in their life, may not enjoy exercise until she becomes more fit and may have conflicting feelings about whether long term benefits are sufficient to outweigh the short term discomfort about exercising. An agent may also be uncertain about the long term benefits compared with a medical practitioner. The other component of β is that the principal may hold more information about the capacity of the agent to perform the task than the agent. This information asymmetry may stem again from the principal's qualifications and experience with the task and through observing a

[8]We make no assumptions about the form of p at this point, as it can represent monetary or non-monetary incentives. However, both β and e can be assumed to belong to $\mathbb{R}_{\geq 0}$.

range of agents undertaking it. For the experiment, an agent may be relatively optimistic or pessimistic in relation to the benefits of undertaking the task, affecting her tendency to put effort into the task. In terms of health, an agent may exhibit time inconsistent behaviour between a stated intention to undertake an exercise programme and her actual effort put into the programme. An agent may have high or low self-belief in relation to her ability to stick with an exercise programme long enough to feel the benefits, compared with her actual ability to successfully develop healthy behaviours. The principal will know how other similar agents have fared with the task previously and therefore can more accurately infer the current agent's ability to undertake the task.[9]

In each time period (or round), after being given the task, the agent infers β from σ (where σ is any prior information they hold about β) and the p chosen by the principal. The principal is uncertain about σ. Therefore, we can assume that the agent will choose effort level $e^*(p, \hat{\beta}(\sigma, p))$, where $\hat{\beta}$ is her conditional expectation of β, given σ and p. To simplify matters, we also assume that the neither the principal nor the agent are aware of future periods, and therefore they are not considering the impact of incentives on future periods.[10]

Thus, to determine the optimal policy setting in any given period, the principal's expected payoff is:

$$E_\sigma[U_P(\beta, e^*(p, \hat{\beta}(\sigma, p)), p)|\beta]. \tag{3.1}$$

Assuming differentiability, the principal's choice of policy is determined by the first order condition:

$$E_\sigma\left[\frac{\partial U_P}{\partial p} + \frac{\partial U_P}{\partial e}\cdot\frac{\partial e^*}{\partial p} + \frac{\partial U_P}{\partial e}\cdot\frac{\partial e^*}{\partial \hat{\beta}}\cdot\frac{\partial \hat{\beta}}{\partial p}\bigg|\beta\right] = 0. \tag{3.2}$$

[9] The term β enters the principal's utility function as we do not rule out the principal caring about the agent's capacity to perform the task. For example, policymakers may care about an individual's capacity to exercise, as well as their level of exercise. This utility function for the principal follows Bénabou and Tirole's (2003) general formulation.

[10] Our experimental design is congruent with this assumption. Subjects are uninformed about future repetitions of the task, as explained in Section 3.3.2. Thus, their σ is affected only by the principal's choice in the previous and current rounds. Hence, we do not complicate matters by deriving the model within a dynamic environment where the principal is concerned about the impact of p on future periods.

Of the three terms in the first order condition, the first corresponds to the direct impact of p on the principal's utility, the second term the direct impact of p on the agent's effort, and the third term on the impact of p on the agent's expectations about β, and thus on their effort. This third term is what Bénabou and Tirole (2003) refer to as the principal's confidence-management motive.

The choice of p will determine how the agent perceives the task and how she perceives the principal's assessment of her capacity to undertake the task. For example, a p that corresponds to a high power monetary incentive may signal to the agent that her cost of undertaking the task is high, due to a low capacity or the principal determining that the agent will not enjoy the task. An alternative and novel example is setting p as a charitable donation. This could signal to the agent that the principal believes in the agent's capacity to undertake the task and that the principal thinks positively of the agent's level of generosity.[11]

In terms of the total effect of p on e^*, a high power financial incentive may raise effort overall, more than compensating for any negative effect of p on $\hat{\beta}$. This effect is behind Bénabou and Tirole's (2003) claim that their model is consistent with the literature on the short term benefits of incentives for effort. However, if p has been removed, e^* may be reduced below the level that it was at before an incentive was applied. Put more precisely, any previous p will impact the information the agent has in the current round, σ, and thus $\hat{\beta}$ is affected in the current round. This result follows from our assumption that the principal and agent are only considering the effect of p on the current round, given all previous rounds, as they are unaware of future rounds.

One application of the model is considering whether to apply an extrinsic incentive to a task that currently relies on intrinsic motivation as the source of effort. An example is exercise. For most individuals, it is their sole responsibility to ensure they exercise enough to stay healthy. In this case the agent is motivating herself, and determining her own capacity to undertake the exercise, β, solely based off her own imperfect observations (which is covered by the term σ in the model). There is no policy intervention; that

[11] For a related application of the Bénabou and Tirole (2003) model to financially incentivising blood donations, see Bolle and Otto (2010).

is, $p = 0$. This thought exercise raises two further questions. First, in the absence of external intervention, is there heterogeneity in σ that can help explain behaviours and life outcomes? Second, what happens when an extrinsic incentive is applied to a task which previously had no extrinsic incentive?

With regards to the first question, there are many potential avenues of exploration. In this paper, we focus on whether there is evidence of an underlying level of intrinsic motivation, or self-drive, that can be observed across domains. Thus, we measure intrinsic motivation to undertake a simple task in the laboratory, and test whether it predicts waist-to-height ratio, which is a good proxy for health outcomes (Ashwell and Hsieh, 2005). This takes intrinsic motivation from one domain, a simple effort task in the laboratory, and applies it to a completely separate domain, health.

The second question determines the underlying design of our experiment. We apply an unexpected extrinsic incentive after initially having none, and then unexpectedly remove it in the next round of the task. This is the within subject element of our design, and it makes our experiment applicable to contexts such as policies to encourage exercise. For the between subject element of our experimental design, we are able to compare the effects a range of incentives. Additionally, this design provides us with the ability to use a differences-in-differences approach to estimating incentive effects, increasing the precision and credibility of our estimates, particularly for the heterogeneous population we sample. Finally, the design also provides us with the ability to investigate the balance between the positive effects of an incentive on effort, and the negative effects of crowding out, shown when the incentive is removed.

3.3.2 Experimental procedures and design

The experiment was run over 12 sessions from 6 April to 3 June, 2016, at the Monash Laboratory for Experimental Economics (MonLEE) at Monash University in Melbourne, Australia. The overall timeline of each experimental session is shown in Table 3.1.

At the start of each session, each subject took a random number from a bucket, which corresponded to one of the 26 computers in the room. They were seated, signed consent

Table 3.1: Overall timeline of each experimental session.

Initialisation	**Activities**	**Surveys**	**Measurement and payment**
Subjects randomly assigned to computers, consent forms signed, overview instructions provided in hard copy and read aloud by experimenter.	Activities for the experiment completed – multiple rounds of a real effort task and a time preferences task (see Table 3.2 for more detail).	Relevant surveys given to subjects.	Subjects instructed to proceed to a neighbouring room to be measured and paid in private by assistants.

forms and then overview instructions were provided in hard copy and read aloud. The instructions outlined the overall session structure, without giving details about the activities themselves. All instructions are included in order in Appendix B.2. The instructions were handed out and read aloud for each round or activity; only the relevant instructions were handed out before each round, with the following round instructions held back until they were needed.

For the initial instructions it was explained to subjects that they may earn money for participating in some of the activities, and that earning details would be explained at the start of the activity. It was also explained that they would be paid at the end of the session by an administrative assistant in a neighbouring room. Next, the activities were undertaken, followed by surveys on health and the experimental activities. These tasks were all undertaken on the computers. When a subject was finished these activities, they were asked to line up outside the neighbouring room where they would have their physical measurements taken and be paid.

The activities section of the experiment proceeded as shown in Table 3.2. We employed multiple rounds of a real effort task in order to address the research aims. We used the word encoding real effort task developed by Erkal et al. (2011), programmed using zTree (Fischbacher, 2007). The number pad on the right-hand side of the keyboard, along with the *Tab* keys, were disabled for all subjects in all sessions to remove the advantage a particularly experienced computer user could have in the task. We use a real effort task as they have been shown to generate utility for subjects and are designed to provide a

Table 3.2: Experimental activities timeline.

Between subject treatment group	**Practice round**	**Round 1**	**Round 2**	**Round 3**	**Time preferences task**	**Round 4**
Control group	Effort task explained and practice round given	Effort task with no incentives; no incentives or future rounds mentioned.	Effort task with no incentives; no incentives or future rounds mentioned.	Effort task with no incentives; no incentives or future rounds mentioned.	Time preferences task explained and given; next effort task round not mentioned.	Effort task with no incentives; no incentives mentioned, not mentioned that this is the last round.
Extrinsic incentive groups (four separate groups, each with a different type of incentive)	As above.	As above.	Effort task with extrinsic incentive (type depending on treatment group); no future rounds mentioned.	Effort task with no incentives; made clear that no incentives are given in this round, no future rounds mentioned.	As above.	Effort task with no incentives; made clear that no incentives are given in this round, not mentioned that this is the last round.
Task time limit (excluding instructions)	2 minutes	5 minutes	5 minutes	5 minutes	No time limit	5 minutes

fine measure of effort on the intensive margin within a short period of time (Brüggen and Strobel, 2007; Gill and Prowse, 2012).

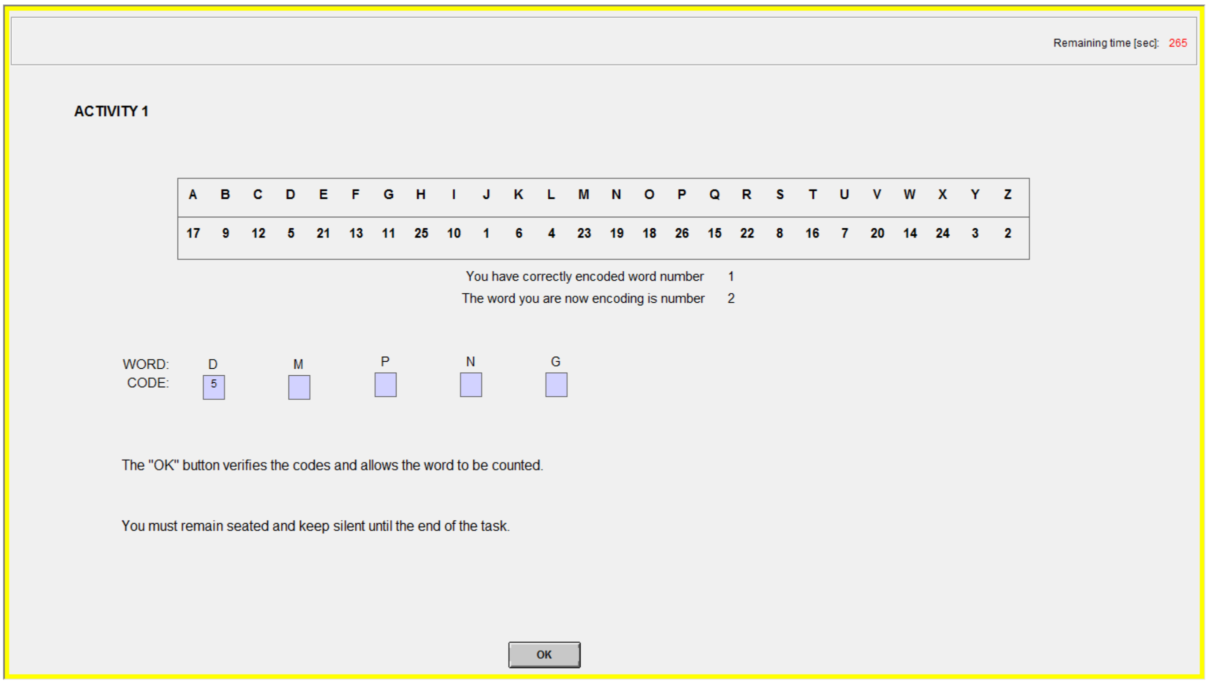
The chosen real effort task is a puzzle, which is a widely favoured type of enjoyable task in the psychology literature (Promberger and Marteau, 2013). Real effort tasks tend to produce positive effort as they provide subjects with utility. The alternative method of a chosen effort task does not provide subjects with such utility (Brüggen and Strobel, 2007) and hence is not appropriate for studying intrinsic motivation. On a related point, we are also clear that we are measuring relative effort on the intensive margin, not the extensive margin as it unlikely that we will fully crowd out effort when there are few outside options and significant motivation to undertake the task (Araujo et al., 2016; Erkal et al., 2017). Indeed, it is the fact that subjects are intrinsically motivated to undertake effort in a real effort task in the laboratory that makes it a suitable choice, as demonstrated by their use in comparable studies (eg. Heyman and Ariely, 2004; Kajackaite and Werner, 2015; Pokorny, 2008).

In the activities portion of the session, first the word encoding task was explained and a 2 minute practice round was given to subjects. An example screenshot of the task is shown in Figure 3.1. The task consists of correctly inputting numbers in the boxes below the 5 randomly selected letters. Once the numbers are correctly inputted and the subject clicks "OK", they are given a new random "word" and a new set of code numbers for each letter of the alphabet. The outcome variable measured from the task is effort in terms of words completed per minute.

After the first round, subjects were told that they would be given the same task again, for another 5 minutes. Those in the control group were given the task as before, without mention of incentives or future rounds of the task.[12] Thus, the control group gives a measure of intrinsic motivation over multiple rounds. However, subjects in the extrinsic incentive treatment groups were given an incentive to complete each word in the task

[12]It is important to note here that there was no deception employed in this experiment. As shown in the overview section of the instructions in the Appendix, subjects were told at the start of each session that "you will be participating in four activities... Each activity may consist of one or more rounds." They were also aware of the general nature of the tasks and the total expected time of the experiment through the sign up process. Subjects were not made aware of the specific details of their next task, as is standard practice in economic experiments. This set up is by design, as explained in this section.

Figure 3.1: Example screen of real effort task given to subjects, with the code for the first letter of the "word" completed.



during this round. The incentive given depended on the between subject treatment group – see Section 3.3.2. Like the control group, the subjects in the incentive groups were not made aware of the future rounds of the task at this stage either. Therefore, round 2 gives a measure of the effect of the incentives, given baseline intrinsic motivation measured in round 1.

After round 2, subjects proceeded to round 3, for which they were given the real effort task for another 5 minutes. There were no incentives given in this round; subjects in the incentive groups were told this fact explicitly, whereas those in the control group were again just given the task without mention of incentives. This round gives a measure of whether the incentives crowd in or crowd out intrinsic motivation, given baseline intrinsic motivation measured in round 1.

Next the subjects were given a time preferences task, which is explained in more detail in Section 3.3.2. This task was given to subjects at this point to test whether the patterns measured in round 3 persist after a break. Thus, after the time preferences task subjects were given the effort task in round 4 for another 5 minutes, with the same treatment as round 3. That is, no incentives were given, which was explained to those in the incentive treatments but not mentioned to those in the control. It was also not mentioned that this was the last round of the effort task.

Payment was received for round 2 (depending on treatment and number of words completed), the time preferences task (between AUD$10 and AUD$20), and AUD$20 for participating in a survey and discrete choice experiment on health during the survey component of the session, which came after round 4 (see Table 3.1). This means subjects earned at least AUD$30 for participating in the session. Payments for the time preference task were made using a gift card (explained more in Section 3.3.2); all other payments were in cash. Given the time required for measurement and payment at the end, subjects were instructed to leave the computer lab and line up outside the neighbouring room once they finished the surveys. This arrangement meant subjects were paid and able to leave between 1 hour and 1 hour and 45 minutes, depending on when they finished the surveys.

Consideration was taken in the experimental design around experimenter demand effects (EDE). Experimenter demand effects are caused by subjects inferring what they are supposed to do in a given situation, which is a particular problem when the EDE are positively correlated with the experimental aims. One of the major causes of EDE stems from the fact that the experimenter is an authority figure, and thus the subjects take cues from the experimenter and the instructions as to how to behave (Zizzo, 2010). In our case, the position of the experimenter is an important part of our theoretical framework and research questions, so this aspect of EDE is not a confound of our results. It does highlight the need to be careful instructions are written in order to frame the real effort task in a way that does not provide an explicit verbal extrinsic incentive to subjects, so that the effort level in round 1 constitutes a good measure of intrinsic motivation. Thus, the instructions for the real effort task were carefully worded to avoid telling subjects to maximise the number of words they completed per minute. Words like “should” and “must” were avoided (see Appendix).

As Zizzo (2010) identifies, framing of instructions is most vital in the case of experiments where the objective of the experimental task is unclear. In our experiment, it is obvious that the purpose of the task is to complete as many words as possible, correctly

Table 3.3: Between subject treatment groups.

Treatment group	**Incentive applied** (in Round 2 only)
Control	None
Low power	\$0.05 paid per word
High power	\$1 paid per word
High power threshold	\$23 paid if complete 23 words; \$1 paid per word above this amount
Charity	Two words plants one indigenous tree within Victoria (equivalent to \$1 per word)

within the time limit. What is not made clear to subjects is that we are testing how motivated they are to put in effort, not just in the first round, but over repeated rounds. We thus use the approach of obfuscation of the experimental objective, including by not making subjects aware of exactly how many rounds they will be completing and with what incentives, to ensure that subjects are not primed to act in a certain way (Zizzo, 2010). Finally, we made it clear at the start that an administration assistant, who is would not be involved the analysing the research data, would be paying the subjects at the end of the session in a private, neighbouring room (see Appendix). This design choice was made to minimise any EDE caused by subjects wanting to please the experimenter by helping to increase the level anonymity of the data collection process, without making it too salient to the subjects to suggest a certain way of acting (Zizzo, 2010).

Between subject treatments

The five between subject treatment groups are shown in Table 3.3. Each session was assigned into one of the treatments. The treatments are defined by the extrinsic incentive applied in round 2 of the activity. As explained in the previous subsection, the control group received no incentive in round 2. The first of the incentive treatments is low power; subjects received AUD\$0.05 for each 5 letter “word” completed during the 5 minute time limit of round 2. The high power treatment group received AUD\$1 per word in round 2. Those subjects in the high power threshold treatment received AUD\$23 if they completed 23 words; below 23 words they received nothing, but for each word completed above 23 they received AUD\$1.

Finally, subjects in the charity treatment were told "every 2 words you complete will fund the planting of one indigenous tree in Victoria. A local environmental charity will receive the funds to plant these trees after the experiment." The charity to which the funds were given, Tree Project, quotes on their website that every AUD$2 donated leads to one tree being planted.[13] Thus, while subjects were not told the monetary amount of their donation until the end of the experiment, it is equivalent to AUD$1 per word completed. The difference from the high power treatment is that donations were made in $2 increments. To ensure credibility of donations, subjects were also told before round 2 started that a session-level donation receipt would be emailed to them to prove the donation had been made, which would include the average number of trees planted per person.

Time preferences task

The time preferences task was given to subjects between rounds 3 and 4 of the real effort task, and provides an important covariate for hypothesis H3 (see Section 3.3.4). Testing hypothesis H3 involves regressing waist-to-height ratio on intrinsic motivation, controlling for various demographics and time preferences. Time preferences are important to include in any regression on waist-to-height ratio as they have been shown tp partially predict choices related to health outcomes (Bradford, 2010; Bradford et al., 2014; Chapman and Coups, 1999). Time preferences could be correlated with intrinsic motivation given the link between time preferences, self control and various life outcomes such as educational performance and wealth (Augenblick et al., 2015; Golsteyn et al., 2014; Moffitt et al., 2011).

The time preferences task consisted of 18 questions as shown in Table 3.4. The nine questions in Table 3.4 were repeated for today versus 5 weeks, and 5 weeks versus 10 weeks. This design means we can determine whether the subject is present or future biased, along with giving us a measure of their level of impatience. All subjects saw all 18 questions.

[13]http://www.treeproject.org.au/, accessed 23 August, 2016.

Table 3.4: Options given in time preferences task questions – for today versus 5 weeks, and 5 weeks versus 10 weeks.

Earlier payment	**Payment 5 weeks later**
$10	$10.05
$10	$10.10
$10	$10.50
$10	$11
$10	$12
$10	$13
$10	$15
$10	$17
$10	$20

It was explained at the start of the task that one question would be randomly selected to be paid out. The payments for this task were made using a gift card that can be used at a variety of stores, including a common supermarket chain and a major department store chain. The gift card was chosen as it can be used at a large number of stores where people commonly shop, and can be sent via the post. It does not have the transaction costs of going to a bank to deposit a cheque.

Our time preferences task design is a modified version of the multiple price list task used by Andreoni et al. (2015). Due to our diverse and non-standard subject pool, we simplify the design to limit the cognitive burden. Our design provides us with blunt measures of impatience and present or future bias.[14] For the former variable we use a count of the number of earlier choices for the 5 weeks versus 10 week payments, and for the latter two we use dummy variables.

3.3.3 Sample

The sample consists of adults over the age of 18. The sample was restricted by not allowing the typical subject pool, undergraduate students, to participate. It was also restricted to

[14]We use only the time preferences task component of the multiple price list task of Andreoni et al. (2015), with further modifications to the price list itself to allow us the ability to hand out $10 gift vouchers at the experiment. The simplified design assumes a linear utility function, which may not be a realistic assumption. Therefore, we use blunt measures of time preferences. Time preference experiments that take into account non-linear utility (eg. Andreoni et al., 2015) are much more complex and therefore cognitively demanding for subjects. For insight into the large literature on how best to measure time preferences see Andreoni and Sprenger (2012) and Miao and Zhong (2015).

those who could travel to Monash University Clayton Campus for one of the scheduled experimental sessions.

Subjects were recruited from Monash University's Centre for Health Economics and the Monash Business Behavioural Laboratory databases, and through other advertisements, including on the Gumtree website (Volunteers Section), the Monash University staff newsletter *The Insider* and the local community newspaper *The Leader*. An example advertisement and email are included in the Appendix. Advertisements were general in nature to not bias the sample towards individuals particularly interested in our research aims; the study was referred to as “an economics experiment aimed at studying behaviours”. The advertisements included some details about the study, such as the weight and waist measurements, to comply with ethics committee stipulations.

It is standard to compensate subjects for their participation in a study such as this one, given the time and travel costs for participation. Thus, a rough figure of earnings for participation was given to potential subjects at the recruitment stage. Abeler and Nosenzo (2015) find that including potential earnings in a recruitment email for a laboratory experiment increases sign up rates threefold compared with no mention of monetary reward, but does not impact the measured prosocial or approval motivations of the subjects. The potential earnings for our experiment were not emphasised to subjects after the initial recruitment stage, and at no point were subjects informed that the amount they earned would be linked to their effort or performance level, until the instructions given at the start of the incentivised round in the experimental activity (round 2). Nevertheless, in our analysis we need to take into account the impact of informing subjects about their potential earnings at the sign up stage; this vague signal is incorporated into the σ parameter of our theoretical model.

Sessions were held on weekdays, at either 12pm or 5:30pm. In order to avoid differences in the composition of the treatment groups, each treatment was assigned to one 12pm session and one 5:30pm session. The aim was to have 50 people in each treatment group. However, the number of no shows in each session had a large variance, meaning it was difficult to reach the required number of subjects in each session. Thus, two smaller extra sessions were run at 12pm for the control and the high power threshold treatments to supplement the numbers in those treatments.

3.3.4 Hypotheses

We test three main sets of hypotheses. For the sake of brevity, we refer to each set of hypotheses as a single hypothesis. We use notation that links the hypotheses to the theoretical framework. Additionally, we define two further terms.

The first term we define for this section is $P = \{p_1, \boldsymbol{p}_2, p_3, p_4\}$, which is the set of extrinsic incentive policies given to subjects over the four rounds. Thus, p_1 is the policy for subjects in round 1. As explained in the preceding section, $p_1 = 0$ for all subjects. Only $\boldsymbol{p}_2$ includes policies other than 0 and therefore it is expressed as a vector of potential policies. The set of policies in $\boldsymbol{p}_2$ have been outlined in Table 3.3.

The second term we add to the notation is T, which is the set of the five treatment groups. The treatment group a subject belongs to is determined by which policy they were given in round 2 (that is, their p_2). Thus, the treatment groups are given the name of the policy applied in round 2, as per Table 3.3.

Our first hypothesis tests the impact of the extrinsic incentives on effort in the round that the extrinsic incentives are applied, round 2. In order to capture the effect of the extrinsic incentives directly, we use a differences-in-differences approach. This approach ensures that we are directly measuring the effect of the incentive on that individual, compared with the control group. Any learning effects for the task will not affect the results.

The null of our first hypothesis is as follows:

H1$_0$: $e^*_{2,c} - e^*_{1,c} = e^*_{2,t} - e^*_{1,t}, \ \forall t \neq c.$

As with the theoretical framework, $e*$ represents effort level given by the individual for a given time period. For the hypotheses, effort level in round 2 for a subject in the control group is given by $e^*_{2,c}$, where the first subscript is the round, and the second subscript is c, which denotes that the individual is in the control treatment group. The term t symbolises treatment group t, which is some treatment group from the set of treatment groups, T.

On average we expect *a priori* that the high power monetary incentives and the charity incentive will be the most effective and thus most likely to show an increase in effort from round 1 to 2, compared with the control group. These predictions follows the findings of DellaVigna and Pope (2017), and Gneezy and Rustichini (2000b) with regards to high power monetary incentives being most effective. DellaVigna and Pope (2017) and Imas (2014) both find effort is insensitive to the size of charity incentives, with Imas (2014) finding charity incentives are more powerful than low power monetary incentives. The low power incentive has the most potential to lower effort in round two, based on the "pay enough or don't pay at all" principal (Gneezy and Rustichini, 2000b), though the opposite could occur if our results are more consistent with Pokorny (2008). The high power incentive with a threshold could have two contradicting effects that may roughly offset each other - raise the effort of those just below the threshold in round 1, whereas those well below the threshold may be discouraged.

Our second hypothesis tests the null that there will be no effect of the incentives on the third and fourth rounds:

H2$_0$: $e^*_{r,c} - e^*_{1,c} = e^*_{r,t} - e^*_{1,t}$, when $r = \{3, 4\}, \forall t \neq c$,

where r is round number.

Consistent with our theoretical model, we predict *a priori* that the monetary incentives will crowd out effort in rounds 3 and 4, with the high power incentives crowding out effort more than the low power incentive. We expect that the charity incentive will crowd in effort, given the previous literature that shows it increases effort when applied (DellaVigna and Pope, 2017; Imas, 2014), and the positive signals a charity incentive sends to the subjects about the nature of the task. We test this hypothesis for the whole sample, and for two subsamples, determined by whether e^*_1 is less than or equal to the median, and whether e_1^* is strictly above the median. The purpose of this further analysis is to investigate whether there is heterogeneity of treatment effects by initial level of intrinsic motivation.

Finally, our third hypothesis tests the relationship between underlying beliefs of individuals in their capabilities, absent any external intervention by the principal, and the health indicator of waist-to-height ratio, w. Stated as the null:

$\mathbf{H3}_0$: If $w = \delta_e e_1^* + \delta_x x + \epsilon$, then $\delta_e = 0$.

In this hypothesis we are testing whether intrinsic motivation predicts waist-to-height ratio, where other relevant covariates are included. The terms δ_e and δ_x are coefficients that detemine the relationship between the covariates and the dependent variable, w. The term x is a vector of demographic variables and ϵ is a random error term, which includes unobserved variables.

This hypothesis tests whether intrinsic motivation is correlated across domains. As discussed in the background section, this hypothesis is motivated by an underlying intrinsic motivation of the individual to apply themselves to a given task at hand. A higher level of this type of intrinsic motivation should be associated with a higher level of effort in the first round, e_1^*, for which no incentive was applied. Thus, we expect *a priori* that a higher level of observed intrinsic motivation will also be associated with better life outcomes; in this case, a lower waist-to-height ratio, w.

We test this hypothesis both for the whole sample, and two subsamples of $w \leq 0.5$ and $w > 0.5$. This subsample split is based on the threshold of waist-to-height ratio suggested by Ashwell and Hsieh (2005), which indicates increased health risks for both men and women, and will allow us to explore whether there is heterogeneity of findings between those with health risks, and those without health risks associated with being overweight.

3.3.5 Analytical approach

We employ a combination of non-parametric and OLS models to test our hypotheses, with a focus on the differences-in-differences between rounds, between treatment groups.

First, we test the differences-in-differences between the first round and rounds 2 through 4, between the treatment groups, to test hypotheses H1 and H2. We estimate the significance of the differences-in-differences using a non-parametric Mann-Whitney U test, and an OLS model. For the OLS model we estimate the equation:

$$e_{i,r} = \alpha + \boldsymbol{\delta}_r \boldsymbol{r}_{i,r} + \boldsymbol{\delta}_t \boldsymbol{t}_i + \boldsymbol{\delta}_{rt}(\boldsymbol{r}_{i,r} \otimes \boldsymbol{t}_i) + \epsilon_{i,r}. \tag{3.3}$$

In this model, $e_{i,r}$ is effort level for individual i in round r, α is the intercept coefficient, and $\boldsymbol{\delta}_r$ is a vector of coefficients on dummy variables for the round, $\boldsymbol{r}_{i,r}$, which are relative to round 1. The vector $\boldsymbol{\delta}_t$ is coefficients on the dummy variables for the treatment group of individual i, $\boldsymbol{t}_i$, which are relative to the control group. Next, $\boldsymbol{\delta}_{rt}$ is a vector of coefficients on the vector of all interactions between rounds and treatment groups, $(\boldsymbol{r}_{i,r} \otimes \boldsymbol{t}_i)$. Finally, $\epsilon_{i,r}$ incorporates the error terms. When estimating this latter model, standard errors are clustered by individual to account for the fact that there are four observations per individual; one from each round, r.

To summarise the coefficients in equation (3.3), the estimates for $\boldsymbol{\delta}_r$ will show any average differences between rounds, $\boldsymbol{\delta}_t$ will show any differences between treatment groups in their effort in round 1 (at which point there were no differences in treatments given to participates), and $\boldsymbol{\delta}_{rt}$ will give the differences-in-differences for rounds 2 to 4, relative to round 1, for each treatment group relative to the control group. It is these latter coefficients that are testing the set of hypotheses in H1 and H2.

Hypothesis H3 is tested by estimating the following equation using OLS:

$$w_i = \alpha + \boldsymbol{\delta}_e \boldsymbol{e}_{i,1} + \boldsymbol{\delta}_x \boldsymbol{x}_i + \epsilon_i. \tag{3.4}$$

This equation follows directly from hypothesis H3. In testing this hypotheses, we include effort level from rounds 2 and 3 in the range of control variables in vector $\boldsymbol{x}_i$. We also note that this equation only has one observation per individual, so there is no need to cluster the standard errors.

3.4 Results

3.4.1 Summary statistics

Table 3.5 summarises the main demographic variables collected on the study subjects, comparing to available data from the 2011 census for Victoria.[15] We do not make any claims of representativeness, but as treatments were randomly assigned we did aim to ensure subjects were similar across the treatment groups.[16] We can also control for the demographic variables in the study analysis. We aimed to have a heterogeneous subject pool for demographic variables such as age, income and waist-to-height ratio. The subject pool is mostly non-students (74%) and entirely non-undergraduate students. Overall, the sample is younger and better educated compared with the census data.

The raw time preference data are shown in Figure 3.2. The figure shows the proportion of subjects choosing the higher future payment, according to the value of that payment. Both the today versus 5 week and 5 week versus 10 week payment choices are shown. While they track each other closely in aggregate, many subjects had different switch-points in the two sets of questions. A different switch-point in the two sets of questions indicates whether the subject is present-biased or future-biased.[17] Of the 230 subjects, 22% show present bias and 28% show future bias.

Table 3.6 shows summary statistics of the health variables collected on subjects. Height, weight, waist and waist-to-height ratio are shown by gender. A waist-to-height ratio of over 0.5 indicates increased health risks, regardless of gender and ethnicity (Ash-

[15]Victoria is the appropriate population of comparison as, according to postcode data collected in the study, some subjects are from parts of Victoria outside of greater Melbourne, even though the study was conducted within Melbourne.

[16]We test for balance between treatment groups on the observables in Table B.1 in the Appendix, and find no evidence of systematic differences.

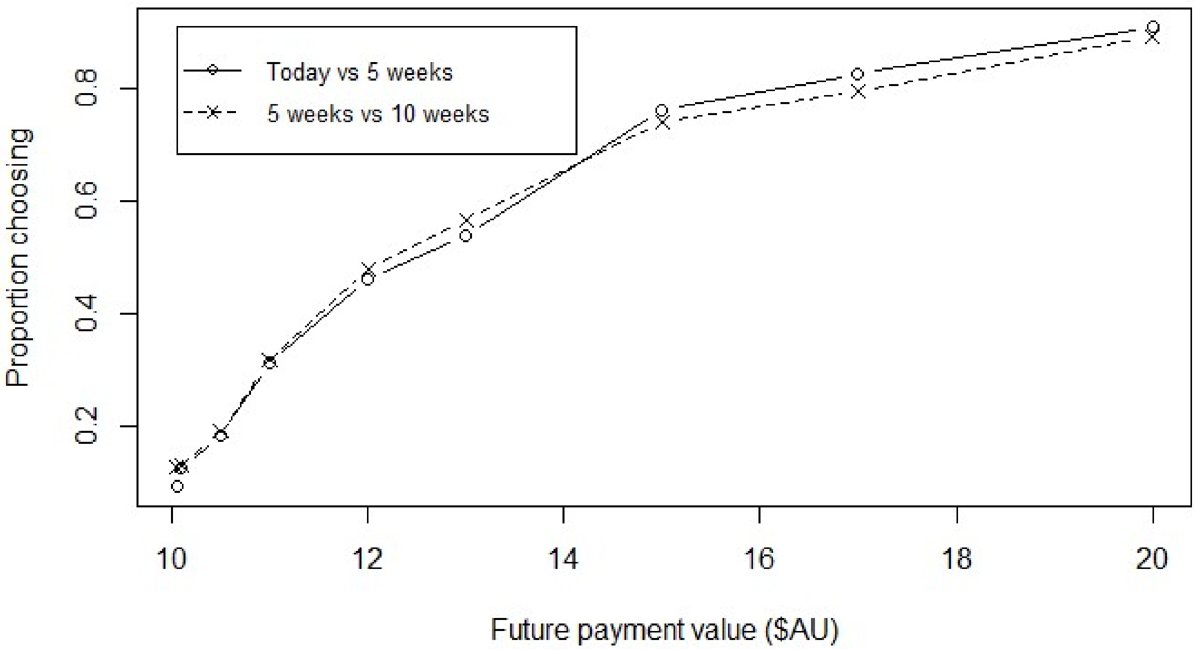
[17]Consider a subject who switches from choosing the earlier payment to the later payment at the $12 mark for today versus 5 weeks, and switches to choosing the later payment at the $11 mark for 5 weeks versus 10 weeks. She will be considered present biased. The opposite case is someone who is future biased.

Table 3.5: Summary statistics - comparing sample demographics to Victoria census data.

	Sample (%)	Census (%)
Gender		
Female	54.8	51.5
Age		
18 to 24	21.3	10.6
25 to 34	33.9	18.3
35 to 44	20.0	18.7
45 to 54	7.4	17.5
55 to 64	8.7	14.7
65 +	8.7	18.4
Education		
Year 11 or other	3.5	34.4
Year 12	10.9	17.9
Certificate	9.1	17.1
Bachelor	46.1	24.2
Graduate	30.4	6.4
Personal income		
Less than $20,000	43.5	
$20,001 to $40,000	24.8	
$40,001 to $60,000	13.9	
$60,001 to $80,000	7.4	
$80,001 to $125,000	8.3	
$125,001 to $150,000	2.2	
Sample/population size	230	4,149,391

Note: Census data from Australian Bureau of Statistics (2011) for those over the age of 18. Census data is only included for data with comparable categories.

Figure 3.2: Time preference choices - proportion choosing higher future payment, over an earlier payment of $10.



well and Hsieh, 2005). Both males and females in our study are just above this threshold on average. The range of the waist-to-height ratio variable is 0.36 to 0.89.

Effort in terms of words completed per minute in each round is summarised in the top half of Table 3.7. There is an overall trend of increasing mean effort from the practice round through to round 4, which is potentially a learning effect. The lowest level of effort ranges from 0 in the practice round and round 4 to 1.4 words per minute in round 2. One subject is dropped from all analysis as the subject did not complete the practice round due to technical issues, lowering the sample size to 229. The subject's round 1 effort was low compared to other rounds, likely due to the lack of a practice round.

Mean effort in round 2 by treatment is shown in the bottom half of Table 3.7. Mean effort is highest in the high power treatment and lowest in the control treatment. This section of the table also shows the number of subjects who undertook each treatment and included in the analysis, which ranges from 44 to 51. The differences in treatment size are due to the high variance in the number of no shows in each session, discussed in Section 3.3.3. Mean effort in each round, by treatment, is shown in Figure 3.3. While the confidence intervals for each round and treatment are overlapping in general, there is one

Table 3.6: Summary statistics of physical measurement variables.

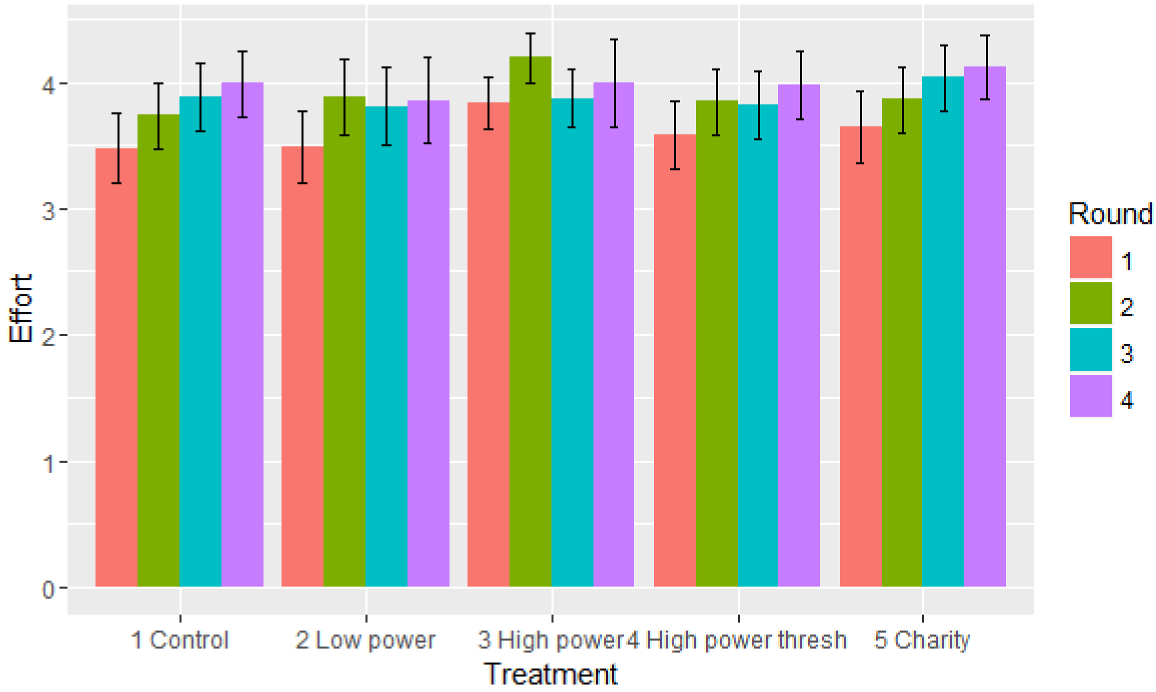
Statistic	N	Mean	St. Dev.	Min	Max
Height (cm)					
Female	126	160.8	6.3	146.0	181.7
Male	104	174.7	7.7	158.0	193.0
Weight (kg)					
Female	126	64.6	15.8	34.6	138.9
Male	104	78.7	19.0	47.8	176.0
Waist (cm)					
Female	126	81.8	14.0	59.0	144.0
Male	104	91.0	13.3	63.5	125.0
Waist-to-height ratio					
Female	126	0.51	0.09	0.38	0.89
Male	104	0.52	0.08	0.36	0.77

Table 3.7: Summary statistics of number of words encoded per minute in each round - pooled sample, and separated by treatment for round 2.

Statistic	N	Mean	St. Dev.	Min	Max
Practice	229	2.7	1.2	0.0	6.0
Round 1	229	3.6	0.9	1.0	6.4
Round 2	229	3.9	0.9	1.4	6.6
Round 3	229	3.9	0.9	0.6	6.6
Round 4	229	4.0	1.0	0.0	6.8
Round 2 by treatment					
Control	44	3.7	0.9	1.8	6.0
Low power	46	3.9	1.0	1.4	6.2
High power	44	4.2	0.7	2.4	5.8
High power threshold	51	3.9	0.9	1.8	5.8
Charity	44	3.9	0.9	2.4	6.6

Note: Between subject treatments differed in Round 2 only.

Figure 3.3: Effort (words per minute) by treatment and round, with 95% confidence intervals.



main trend worth noting. Effort of those in the control and charity treatments increases each round, whereas this is not the case for those in the monetary incentive treatments (low power, high power and high power threshold). In these three treatments, effort is increasing between each round except for rounds 2 to 3, where there is a decrease.

3.4.2 Effects of extrinsic incentives

In this section we present the main econometric results relating to hypotheses H1 and H2, followed by our findings on these hypotheses. Hypothesis H1 relates to the effect of extrinsic incentives when they applied in round 2, and H2 relates to the effect of removing the incentives on effort in rounds 3 and 4. In Table 3.8 we present p-values from the non-parametric Mann-Whitney U test for the difference in effort between rounds 1 and rounds 2 to 4, for each treatment compared with control.

In the first row of results in Table 3.8 we aggregate all groups with an extrinsic incentive. The first result in the top left of the table shows there is no statistical difference in the differences between effort in round 1 and round 2 between groups with an incentive

Table 3.8: Mann-Whitney U test p-values for differences-in-differences of effort between each treatment group and control, between round 1 and rounds 2 to 4.

	Differences-in-differences compared with control: p-value		
Treatment	R1 to R2	R1 to R3	R1 to R4
Aggregated treatment groups			
All incentives	0.351	0.093*	0.222
Monetary incentives	0.138	0.037**	0.179
Disaggregated treatment groups			
Low power	0.065*	0.426	0.352
High power	0.331	0.001***	0.071*
High power thresh	0.389	0.235	0.530
Charity	0.452	0.882	0.588

Notes: R1 is shorthand for round 1, etc. *p<0.1; **p<0.05; ***p<0.01.

in round 2 and with no incentive in round 2 (the control group). Moving rightwards, the second result shows there is a statistical difference between round 1 and round 3 effort between those with an incentive and those without, at the 10% level. This result does not hold for the differences in round 1 and round 4. The second row of results repeats the exercise, but comparing just the monetary incentive treatment groups (that is, all incentivised groups except those with the charity incentive) with the control. The results show the same pattern, but all have lower p-values.

Moving to results in Table 3.8 for the disaggregated treatment groups, there are three statistically significant differences-in-differences. First is the difference in effort between rounds 1 and 2 for the low power treatment, at the 10% level. Second is the difference in effort between rounds 1 and 3 for the high power treatment, at the 1% level, which is the most statistically significant result in the table. Third is the difference in effort between rounds 1 and 4 for the high power treatment, at the 10% level.

Next we estimate a differences-in-differences model using OLS, as per equation (3.3) in Section 3.3.5. The results are shown in Table 3.9. Column (1) of the table present the estimates for the full sample, therefore we discuss these first.

After the constant term, the first three variables in Table 3.9 are round dummies,

Table 3.9: Differences-in-differences models of rounds 1 to 4, including all treatments.

	Dependent variable:		
	Rounds 1 to 4, words/minute		
	All	R1 > median	R1 ≤ median
	(1)	(2)	(3)
Constant	3.477***	4.242***	2.896***
	(0.138)	(0.127)	(0.136)
Round 2	0.264***	0.242***	0.280***
	(0.054)	(0.071)	(0.080)
Round 3	0.414***	0.379***	0.440***
	(0.070)	(0.096)	(0.102)
Round 4	0.518***	0.453***	0.568***
	(0.068)	(0.094)	(0.097)
Low power	0.010	0.113	0.033
	(0.198)	(0.198)	(0.186)
High power	0.368**	0.012	0.360**
	(0.171)	(0.161)	(0.162)
High power thresh	0.103	0.316*	0.104
	(0.193)	(0.183)	(0.173)
Charity	0.168	0.247	0.166
	(0.197)	(0.198)	(0.179)
Low power*R2	0.141*	0.147	0.134
	(0.076)	(0.101)	(0.110)
High power*R2	0.091	−0.027	0.276**
	(0.083)	(0.095)	(0.134)
High power thresh*R2	0.007	−0.189	0.120
	(0.080)	(0.134)	(0.096)
Charity*R2	−0.041	−0.142	0.028
	(0.088)	(0.101)	(0.132)
Low power*R3	−0.088	−0.001	−0.147
	(0.100)	(0.127)	(0.146)
High power*R3	−0.382***	−0.517***	−0.162
	(0.122)	(0.150)	(0.192)
High power thresh*R3	−0.163*	−0.284*	−0.096
	(0.093)	(0.149)	(0.121)
Charity*R3	−0.018	−0.146	0.068
	(0.108)	(0.137)	(0.156)
Low power*R4	−0.144	0.025	−0.261
	(0.115)	(0.134)	(0.169)
High power*R4	−0.364**	−0.360*	−0.324
	(0.171)	(0.218)	(0.281)
High power thresh*R4	−0.118	−0.305*	−0.018
	(0.098)	(0.171)	(0.115)
Charity*R4	−0.036	−0.130	0.024
	(0.103)	(0.132)	(0.148)
N	229	100	129
Observations	916	400	516
Adjusted R^2	0.019	0.031	0.077
F Statistic	13.576***	6.0501***	11.835***

Notes: Standard errors clustered at the individual level are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

relative to round 1. They show that the dependent variable, effort (words per minute) increases in each round, controlling for treatment group. This would appear to be a learning effect, and is consistent with the summary statistics shown in Table 3.7 and Figure 3.3.[18]

The next four variables in descending order in Table 3.9 are dummies for each treatment. These coefficients pick up whether there is any difference between the incentive treatment groups and the control group effort in round 1. The high power incentive treatment group appears to have a higher level of mean effort overall, while all the other treatment groups do not have a different mean effort level from the control group.[19]

The four variables following are treatment dummies interacted with the round 2 dummy, which show the impact of the incentive treatments relative to the control group in round 2. Only the low power incentive leads to statistically more effort than the control group in round 2. This finding is consistent with the non-parametric testing shown in Table 3.8. The coefficient on the low power incentive in round 2 can be interpreted as an increase of 0.141 words per minute of effort, which is equivalent to a 3.8% increase in effort in round 2 over the control group. However, it should be noted that the high power incentive has a positive coefficient of 0.091. Thus, while the coefficient is not statistically significant, it still is consistent with the high power incentive having a positive effect within a similar range as the low power incentive. Furthermore, the initially higher effort of those in the high power treatment group in round 1 may account for the lower difference-in-differences of round 1 to round 2 effort for the high power treatment group compared with the low power treatment group.

The incentive treatment-round 3 interaction dummy variables show a crowding out of effort in the high power and high power threshold incentive treatments. The high power incentive crowding out is particularly large, amounting to 9.8% less effort in the high power treatment group compared with the control group in round 3. Crowding out of

[18] The inclusion of these round dummies ensures this learning effect is accounted for when testing the hypotheses.

[19] As already mentioned in a previous footnote, Table B.1 in the appendix shows this difference in average effort level between the high power treatment group and the control group cannot be attributed to any major differences in the observables.

effort is not present in the low power or charity treatment groups, nor is crowding in.

There are similar results between round 3 and round 4, though the effects observed in round 3 are partially dissipated by round 4. Round 4 is designed to test whether any effects observed in round 3 persist, given the break provided by a different task between rounds 3 and 4. Only the high power incentive treatment is statistically different from the control group in round 4, but is almost as far below the control group as the high power group in round 3.

Columns (2) and (3) in Table 3.9 show the results from the same model as column (1), but for subsamples with effort in round 1 above median (column (2)) and equal to or below median (column (3)). For both these subsamples, the low power incentive no longer provides enough of an incentive to increase effort in round 2 to show up statistically, though the coefficients for both subsample are still very similar to column (1). However, the low motivation subsample in column (3) is positively motivated in round 2 by the high power incentive at the 5% level of significance.

In terms of crowding out effects in rounds 3 and 4, the high effort subsample in column (2) has significant crowding out from both high power incentives in both rounds 3 and 4. There is however no statistically significant crowding out for the low motivation subsample in column (3).

Effect of extrinsic incentives during application

The non-parametric and parametric models presented in Tables 3.8 and 3.9 both support the following finding in relation to hypothesis H1:

Result 1: Only the low power incentive has a statistically significant positive impact on effort for the full sample, including all treatment groups. For all other incentives we fail to reject the null hypothesis $H1_0$, that the incentives have no effect on round 2 effort.

The result for the low power incentive is significant at the 10% level. It is worth noting that this result is found using the differences-in-differences. A straight comparison of effort levels in round 2 finds that the high power incentive is effective at raising effort at the 1% level, but this finding does not take into account the higher effort level of the high power treatment group in round 1, as shown in Table 3.9. This example shows the strength of our experimental design, which allows us to use the differences-in-differences method to analyse the effect of incentives.

From columns (2) and (3) in Table 3.9, we also find the following result:

Result 1a: The high power incentive is effective at raising effort for individuals at or below the median level of intrinsic motivation observed in round 1.

This result holds at the 5% level, even though it is found using just over half the total sample size. No other incentives have a statistically significant effect on either of the two subsamples used in columns (2) and (3) of Table 3.9.

Effect of extrinsic incentives after removal

From Table 3.9 we present the following result, related to hypothesis H2:

Result 2: The high power incentive has a statistically significant negative impact on effort after it is removed, for rounds 3 and 4. The high power threshold incentive has a statistically significant negative impact on effort in round 3. For all other incentives we fail to reject the null hypothesis $H2_0$, that the incentives have no effect after their removal for rounds 3 and 4.

This result is consistent with the high power incentives crowding out intrinsic motivation. It is supported by the non-parametric testing in Table 3.8 for the high power incentive, but not for the high power threshold incentive. However, Table 3.8 does show that all incentives, and aggregated monetary incentives have a statistically significant effect on effort in round 3 compared with the control; this effect is negative. The effect does not persist to round 4.

From the subsample models of columns (2) and (3) in Table 3.9, we also find:

Result 2a: The high power and high power threshold incentives have a statistically significant negative effect after their removal, for rounds 3 and 4, for individuals with intrinsic motivation observed in round 1 above the median. There is no crowding out effect observed for individuals with intrinsic motivation at or below the median.

The observed crowding out from the high power incentives for high intrinsic motivation individuals is at the 10% level, except for the high power incentive in round 3, which is at the 1% level. In this latter case the crowding out effect represents an 11.2% decrease in effort in Round 3 compared with the control group.

3.4.3 Intrinsic motivation and health

In this section we test whether our measure of intrinsic motivation, effort in round 1, provides additional explanatory power for waist-to-height ratio, a measure of individual health. Thus, we are testing our third hypothesis. The results are shown in Table 3.10.

In column (1) of Table 3.10 we regress waist-to-height ratio on just effort level in round 1. There is a negative relationship between the two variables at the 1% level. Column (2) repeats the same exercise, but with age being the only covariate in the model. There is a positive relationship with age at the 1% level. Column (3) includes both age and round 1 effort. Both coefficients maintain their sign and level of statistical significance, but both are reduced in absolute value. The largest effect is on round 1 effort, with the coefficient roughly halving, whereas the coefficient on age has a more modest decrease.

Column (4) of Table 3.10 regresses age and the other demographic variables on waist-to-height ratio, dropping round 1 effort. The coefficient on age is marginally smaller than column (2), but larger than in column (3). There is no statistically significant coefficient estimated on gender, as expected given one advantage of waist-to-height ratio as a measure of health risk is that it is robust to gender and ethnicity (Ashwell and Hsieh, 2005). There is a statistically negative relationship found between years of education and waist-to-height ratio. Personal income does not have any effect, whereas increased impatience is associated with a higher ratio.

We add round 1 effort in column (5), keeping all other variables from column (4). The coefficient on round 1 effort is again reduced in absolute value, but maintains a strong statistical significance - at the 5% level. The significant coefficients on age, education and impatience all lower in size.

Thus, we find the following result in relation to hypothesis H3:

Result 3: Intrinsic motivation, measured through effort level in round 1, has a negative relationship with waist-to-height ratio at the 5% level. Thus, we reject the null hypothesis $H3_0$, that intrinsic motivation does not have explanatory power for waist-to-height ratio.

This result is robust to important demographic controls, including age, education, income and impatience. It is in line with the alternative hypothesis that intrinsic motivation is associated with better personal maintenance of health.

We look at whether this result is heterogeneous in the sample in columns (6) and (7), dividing the sample between those with a waist-to-height ratio of over 0.5, and of less than or equal to 0.5. This delineation is on the threshold value that predicts poor health outcomes (Ashwell and Hsieh, 2005), and coincidentally also splits the sample into two roughly equal sized subsamples. Two main findings stand out from this exercise.

First, waist-to-height ratio is only predicted by intrinsic motivation and demographic variables for those with a waist-to-height ratio above 0.5. Second, age does not predict waist-to-height ratio for either subsample, and therefore must predict membership into the subsamples but not the distribution within them. These findings lead to the following result:

Result 3a: Result 3 is driven by individuals with a waist-to-height ratio above 0.5.

Table 3.10: Waist-to-height ratio regressed on round 1 effort and demographic variables.

	Dependent variable:						
	Waist-height ratio						
	All					WHR > 0.5	WHR ≤ 0.5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.6421***	0.4176***	0.4984***	0.5333***	0.5994***	0.7733***	0.4157***
	(0.0210)	(0.0125)	(0.0312)	(0.0532)	(0.0594)	(0.0702)	(0.0385)
Effort R1	−0.0355***		−0.0172***		−0.0151**	−0.0157**	−0.0042
	(0.0057)		(0.0061)		(0.0063)	(0.0073)	(0.0041)
Age		0.0026***	0.0021***	0.0024***	0.0019***	0.0003	0.0000
		(0.0003)	(0.0004)	(0.0003)	(0.0004)	(0.0004)	(0.0003)
Female				−0.0082	−0.0061	0.0116	−0.0007
				(0.0096)	(0.0095)	(0.0113)	(0.0059)
Education (years)				−0.0091***	−0.0087***	−0.0133***	0.0028
				(0.0033)	(0.0033)	(0.0036)	(0.0023)
Personal income ($1K)				0.0002	0.0003	0.0004*	0.0001
				(0.0002)	(0.0002)	(0.0002)	(0.0001)
Impatience				0.0042**	0.0035*	0.0047**	0.0008
				(0.0019)	(0.0019)	(0.0024)	(0.0011)
Present bias				0.0167	0.0139	−0.0170	0.0086
				(0.0123)	(0.0122)	(0.0150)	(0.0075)
Future bias				0.0006	0.0022	−0.0050	−0.0016
				(0.0116)	(0.0115)	(0.0131)	(0.0070)
N	229	229	229	229	229	112	117
Observations	229	229	229	229	229	112	117
Adjusted R^2	0.1442	0.2333	0.2561	0.2645	0.2799	0.2625	−0.0098
F Statistic	39.4139***	70.3881***	40.2456***	12.7159***	12.0797***	5.9375***	0.8595

Notes: Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 3.11: Waist-to-height ratio regressed on effort in rounds 1 to 3 and demographic variables.

	Dependent variable:						
	Waist-height ratio						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.6421***	0.6534***	0.6254***	0.6541***	0.6413***	0.5742***	0.5916***
	(0.0210)	(0.0232)	(0.0229)	(0.0231)	(0.0227)	(0.0605)	(0.0607)
Effort R1	−0.0355***			−0.0213*	−0.0363***		−0.0200**
	(0.0057)			(0.0127)	(0.0101)		(0.0098)
Effort R2		−0.0356***		−0.0162			
		(0.0058)		(0.0130)			
Effort R3			−0.0286***		0.0009	−0.0084	0.0060
			(0.0057)		(0.0099)	(0.0060)	(0.0092)
Age						0.0021***	0.0019***
						(0.0004)	(0.0004)
Female						−0.0071	−0.0062
						(0.0096)	(0.0095)
Education (years)						−0.0091***	−0.0086***
						(0.0033)	(0.0033)
Personal income ($1K)						0.0002	0.0003
						(0.0002)	(0.0002)
Impatience						0.0040**	0.0034*
						(0.0019)	(0.0019)
Present bias						0.0158	0.0137
						(0.0123)	(0.0123)
Future bias						0.0012	0.0023
						(0.0115)	(0.0115)
N	229	229	229	229	229	229	229
Observations	229	229	229	229	229	229	229
Adjusted R^2	0.1442	0.1395	0.0952	0.1462	0.1404	0.2678	0.2780
F Statistic	39.4139***	37.9536***	24.9995***	20.5276***	19.6251***	11.4217***	10.7562***

Notes: Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

We further investigate our measure of intrinsic motivation, plus the impacts of incentives in Table 3.11. Here we estimate how well effort level in rounds 1 to 3 predict waist-to-height ratio. In columns (1) to (3) we regress effort in rounds 1 to 3 individually on waist-to-height ratio.

Column (1) in Table 3.11 repeats the model in column (1) of Table 3.10, showing the negative and highly statistically significant predictive power of effort in round 1 for waist-to-height ratio. Waist-to-height ratio regressed only on effort in round 2 is shown in column (2). The coefficient on round 2 effort is almost identical to the coefficient on round 1 effort and is again highly statistically significant. Column (3) shows the coefficient for round 3 effort, the round for which the incentives were removed. This coefficient is also negative but smaller in absolute value compared with the coefficients on rounds 1 and 2, but is also statistically significant at the 1% level. We leave out round 4 from the table as the overall findings for round 4 are similar to round 3 and thus it does not add much to the analysis.

In order to test the relative predictive power of each round, we regress waist-to-height ratio on both rounds 1 and 2 in column (4). Using both rounds decreases the absolute value of the two coefficients, as would be expected given the two variables are highly correlated ($r = 0.90$). However, round 1 effort proves to be a stronger predictive variable for waist-to-height ratio, with a larger coefficient (in absolute terms). Additionally, the coefficient on round 1 effort is statistically significant at the 10% level, compared with no statistical significance for the coefficient on round 2 effort.

The same exercise from column (4) is repeated in column (5), but this time with round 1 and round 3. Rounds 1 and 3 are also highly correlated ($r = 0.83$), but only round 1 effort has any predictive power for waist-to-height ratio, with a highly statistically significant negative coefficient. Thus, it appears adding then removing incentives has crowded out the intrinsic motivation that is predictive of waist-to-height ratio in round 1 effort but not in round 3.

To test this finding further, we regress waist-to-height ratio on round 3 effort and the demographic variables in column (6). Round 3 effort has no statistical predictive powers for waist-to-height ratio once these controls are added in. Finally, we also add round 1 effort in column (7), along with round 3 effort and demographic variables. Round 3 effort remains with no statistical significance, and round 1 effort maintains a strong predictive power for waist-to-height ratio, albeit with a smaller coefficient as found in column (5) of Table 3.10. Hence, we find:

Result 3b: Removing the temporary incentives has the effect of removing the intrinsic motivation for effort that is predictive of waist-to-height ratio.

We break this result down by treatment group in Table 3.12 to investigate how this effect operates across the control and treatment grousp individually. Column (1) shows waist-to-height ratio regressed on effort round 1 for the control group, again showing the strong negative relationship between the two variables. Column (2) adds effort round 3 to the regression. The coefficients for effort rounds 1 and 3 are not individually significant, but jointly significant at the 1% level, showing the high level of correlation between the two. The same exercise is repeated for the low power monetary incentive in columns (3) and (4). Effort round 1 is strongly predictive of waist-to-height ratio in column (3), but when effort in round 3 is added (column (4)), the coefficient for effort round 1 is still statistically significant, and the coefficient for effort round 3 is not. This pattern is repeated for all other incentive treatment groups, shown in the remaining columns.[20]

This result has two major implications. First, it suggests that there may be some crowding out of intrinsic motivation by all incentive treatments for round 3, which are too small to be detected in our Table 3.9 given our sample size. Indeed, all treatment-round 3 interaction coefficients for the full sample are negative in that table, even if they are not all statistically significant. Second, it reinforces that our results are highly applicable to policy, particularly to policies to encourage healthy behaviours. We discuss overall implications of our results in more detail in the following section.

[20]The exception is the charity treatment group in columns (9) and (10), but the significance of the coefficient on effort round 1 is low by itself in column (9). In column (10), the coefficient on effort round 1 is almost unchanged, the coefficient estimate for effort round 3 is close to zero, and the model fit is worsened. Thus, it is safe to conclude that the pattern is consistent with the other incentive treatments.

Table 3.12: Waist-to-height ratio regressed on effort in rounds 1 and 3 by treatment group subsample.

	Dependent variable: Waist-height ratio									
	Control		Low power		High power		High power thresh.		Charity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.6787***	0.7064***	0.6394***	0.6275***	0.7611***	0.7706***	0.6173***	0.6056***	0.5877***	0.5823***
	(0.0450)	(0.0511)	(0.0374)	(0.0380)	(0.0777)	(0.0832)	(0.0420)	(0.0448)	(0.0473)	(0.0566)
Effort R1	−0.0461***	−0.0215	−0.0338***	−0.0619***	−0.0622***	−0.0571**	−0.0318***	−0.0496*	−0.0223*	−0.0255
	(0.0125)	(0.0251)	(0.0104)	(0.0224)	(0.0199)	(0.0250)	(0.0113)	(0.0259)	(0.0126)	(0.0222)
Effort R3		−0.0291		0.0289		−0.0075		0.0197		0.0043
		(0.0258)		(0.0204)		(0.0217)		(0.0257)		(0.0239)
N	44	44	46	46	44	44	51	51	44	44
Observations	44	44	46	46	44	44	51	51	44	44
Adjusted R^2	0.2264	0.2314	0.1763	0.1947	0.1691	0.1513	0.1212	0.1137	0.0473	0.0248
F Statistic	13.5820***	7.4720***	10.6321***	6.4414***	9.7523***	4.8339**	7.8939***	4.2064**	3.1337*	1.5470

Notes: Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

3.5 Discussion

Our three sets of results provide insights on intrinsic motivation, extrinsic incentives and the relevance of our laboratory findings for the field. In this section, we discuss each of these sets of results, our methodological contribution and the policy implications.

The overall finding from Result 1 is that only the low power incentive raises the effort of subjects, on average for the full sample. This result is found using both non-parametric and parametric methods. Given the statistical significance of this finding is at the 10% level, our results suggest a high overall level of intrinsic motivation is present among subjects, which is associated with a low level of responsiveness to incentives. The low power incentive raises effort by 3.8%, compared with no incentive. This effect is small, but not inconsequential. Thus, we find support for Pokorny (2008), to go small with monetary incentives.[21] We find this result only at the pooled level for the sample data.

Result 2 is a strong finding, that the high power incentive crowds out effort after its removal, in rounds 3 and 4, for the full sample. This crowding out is significant at the 1% level in round 3, using parametric and non-parametric testing, and is a sizeable 9.8%. Crowding out persists for round 4, but at a diminished rate. There is no statistically significant crowding out found for the low power incentive. These findings suggest that the low power incentive raises effort when it is applied in round 2 because there is no significant crowding out effect. The high power incentive does not successfully raise effort in round 2 as its positive effect on effort is counteracted by its crowding out effect. The crowding out effect remains for round 3, after the incentive is removed.

Results 1a and 2a show considerable heterogeneity in subjects' responses to the treatments. Subjects with a higher measure of intrinsic motivation (that is, effort above the median in round 1) do not show a statistically significant change in effort in round 2 from any of the incentives. They do, however, demonstrate a large crowding effect from the high power incentives in rounds 3 and 4 (of 11.2% and 7.7% respectively in the case of the

[21]Even recognising that we cannot rule out that the high power incentive has a similar effect size to the low power incentive, our findings still support Pokorny (2008) on two grounds. First, why would a principal pay a high incentive, when he can get a similar effect from a low incentive? Second, the high power incentive has a strong crowding out effect whereas the low does not, as discussed next.

pure high power incentive). In contrast, the low motivation individuals are responsive to high power incentives at the 5% level of significance (an 8.7% increase in effort), and do not show any statistically significant crowding out once incentives have been removed in rounds 3 and 4.

Overall, Results 1a and 2a suggest that a higher level of intrinsic motivation leads to a lower responsiveness to incentives, as the positive effects of those incentives are associated with an overall higher level of crowding out among this more motivated group. Individuals with lower levels of intrinsic motivation are most responsive to high power incentives as they do not experience significant crowding from such incentives. These findings are consistent with our overall contention about how intrinsic motivation operates. When an incentive crowds out intrinsic motivation, it may not raise individual effort when it is applied because the crowding out effect roughly balances the effect of the incentive. Once removed, only the crowding out effect remains and effort is reduced. When the incentive does not crowd out effort significantly, it may raise effort when it is applied, and not significantly reduce effort on its removal.

Finally, in relation to these first two main results, it is worth noting that the charity incentive in this case is not powerful enough to increase effort. This is in contrast to the findings of Imas (2014) and DellaVigna and Pope (2017). Building on these studies though, with our experimental design we can also see that the charity incentive also does not have a significant crowding out effect. For the high power threshold incentive, the threshold effectively reduces the power of the high power incentive, both for increasing effort in round 2 and in relation to the level of crowding out. This finding is shown throughout the results in Table 3.9. For example, the low motivation subjects are responsive to the high power incentive when applied in round 2, but not for the high power threshold incentive. The threshold for this incentive (below which subjects did not earn anything) was set above the average level of effort, so for many subject perhaps it was seen as no incentive as they did not believe they could reach that threshold. In future experiments it is worth testing how subjects would respond to an individually tailored threshold – for example, one that is 10% higher than their effort in the previous round.

Result 3 demonstrates that underlying intrinsic motivation has the potential to explain real world outcomes, which in our case is health outcomes. Even after adding important covariates (including age, gender, education, income and time preferences), intrinsic motivation, measured by effort in round 1, still has a negative relationship with waist-to-height ratio at the 5% level of significance. As found with Result 3a, these results are driven by those with a waist-to-height ratio above 0.5, which indicates heightened health risk (Ashwell and Hsieh, 2005). Thus, it seems that lower intrinsic motivation is associated with worsening health risks, for those already above the risky weight.[22]

Result 3b strengthens these findings. As with effort level in round 1, round 3 effort is highly correlated with waist-to-height ratio and predicts waist-to-height ratio at the 1% level of significant when it is the only covariate in the regression. However, round 3 no longer predicts waist-to-height ratio when the full set of covariates are added to the regression. This finding suggests the explanatory power that intrinsic motivation has for waist-to-height ratio has been crowded out for those individuals receiving extrinsic incentives. Including effort in round 3 actually marginally increases the absolute size of the coefficient on round 1 effort level, as it may be controlling for a small component of underlying ability in the task that is correlated with waist-to-height ratio and is not controlled for by the other covariates.

The departure of our experimental design from the common one round, between subject, comparison of incentives has allowed us to provide several new insights. Our multi-round, within and between subject experimental design allows us to control for baseline intrinsic motivation, providing us with increased statistical power and more accurate results on the power of incentives. For example, we can control for the fact that the high power incentive treatment group has statistically higher baseline effort. The design allows us to test for a crowding out effect not only during the application of incentives, but also

[22]We additionally test what predicts whether an individual has a waist-to-height ratio above 0.5, to determine whether intrinsic motivation has a role in determining whether or not individuals are in the risky weight category. Age is the strongest predictor of having a waist-to-height ratio above 0.5, at the 1% level, with the other significant coefficient being a negative coefficient on the female dummy at the 10% level. Intrinsic motivation does not predict whether or not an individual has a waist-to-height ratio above 0.5 ($p = 0.20$), when the other covariates are included. These results are found both with an OLS and a probit model.

after their removal. This feature of the design helps us better understand the balance between the positive effects of incentives, and the negative effects of crowding out. Additionally, our measure of baseline intrinsic motivation has allowed us not only to uncover heterogeneous effects of incentives, but also the result that intrinsic motivation can explain some of the variation in health outcomes. Finally, we use an objective measure of health risk (waist-to-height ratio) to apply the laboratory data to outcomes in the field. This approach avoids some of the criticisms of using survey measures, which are more typically used in a laboratory, such as that subjects may not be able to accurately recall their general level of health.

Results 3 to 3b provide evidence of a positive relationship between intrinsic motivation and health outcomes, and show that all extrinsic incentives have the potential to crowd out intrinsic motivation. Thus, we suggest caution should be applied when considering the incentivisation of health behaviours. Our study does highlight the importance of understanding heterogeneity when it comes to intrinsic motivation and extrinsic incentives. We find small monetary incentives are most effective at the population level for increasing effort, but large monetary incentives are more effective for low motivation individuals.

Our findings suggest future research into targeting incentives at low motivation individuals. This research should investigate what effect common knowledge of this targeting has on both the low and high effort individuals, given our subjects have common knowledge that all other subjects in their session face the same incentives. Research on common knowledge is important when considering policy interventions that are openly targeted at specific individuals, as these policies may increase crowding out effects for those individuals identified as low performing.

3.6 Conclusion

In this paper we present a lab-in-the-field experiment on intrinsic motivation, its importance in explaining behaviours, and show some conditions under which extrinsic incentives can crowd out intrinsic motivation. We employ a rich within and between subject design that allows us to use a differences-in-differences approach to test the main hypotheses.

Additionally, we apply our measure of intrinsic motivation to health outcomes, using the objective measure of waist-to-height ratio. Our study combines a diverse subject pool with an experimental design that allows us to both control for and more deeply understand heterogeneity.

For the full sample, we find support for “pay – but do not pay too much” (Pokorny, 2008). However, for low motivation individuals, we find “pay enough or don’t pay at all” (Gneezy and Rustichini, 2000b) is a better rule to follow. Finally, we find that intrinsic motivation is an omitted variable for explaining health outcomes, in particular waist-to-height ratio. Given that temporary extrinsic incentives can crowd out intrinsic motivation, incentives aimed at encouraging healthy behaviours need to be considered with caution. Crowding out effects are shown to persist, but at a diminishing rate over time, which provides an impetus for future research into the longevity of crowding out effects. The application of our lab-in-the-field data to health outcomes demonstrates the value of measuring intrinsic motivation within a laboratory context for policy development and for applied research in the field.

Chapter 4

A behavioural rebound effect: Results from a laboratory experiment

4.1 Introduction

Behavioural "nudges" provide a powerful avenue for decreasing energy use (Allcott and Mullainathan, 2010). Environmental campaigners, corporations, governments and economists all understand that individuals have pro-environmental preferences and a proclivity to follow social norms, both of which can lead to pro-environmental choices and behaviours (Cason and Gangadharan, 2002; Croson and Treich, 2014; DEFRA, 2008). At the same time, technological change also forms a vital part of environmental policy, for everything from addressing water shortages to climate change (Duarte et al., 2014; Global Commission on the Economy and Climate, 2014). But a change in technology alters incentives. For example, moving to a more efficient car decreases the relative environmental benefit of walking and cycling, thus reducing the pro-environmental incentives for not driving. This leads to the question – what might be the behavioural effects resulting from a change in technology?

An increase in consumption due an increase in energy efficiency, or rebound effect, has long been recognised (Jevons, 1865). The rebound effect is modelled in the literature as a result of simple income and substitution effects (Chan and Gillingham, 2015). While private income and substitution effects are clearly important drivers of choices, a behavioural rebound effect related to changes in pro-environmental incentives has yet to be explored.

The aim of this paper is to investigate the existence of a behavioural rebound effect and whether improvements in energy efficiency are subject to moral licensing. In this paper I define the behavioural rebound effect as a decrease in pro-environmental effort after an increase in energy efficiency. Pro-environmental effort refers to effort undertaken purely for environmental reasons, such as any walking or cycling done purely on environmental grounds and not for other benefits from these modes of transport, like fitness, enjoyment or saving money. As with the standard rebound effect, the behavioural rebound effect is defined in relation to an exogenous change in energy efficiency. Moral licensing accounts for any additional reduction in pro-environmental effort due to an endogenous change in energy efficiency. Moral licensing is a behavioural phenomenon whereby individuals who undertake a moral action will subsequently behave in an immoral or unethical way (Blanken et al., 2015); Tiefenbeck et al. (2013) find evidence for moral licensing within the domain of household water and energy consumption. The aforementioned transport example could also include a moral licensing effect. After an individual purchases a highly efficient car at least ostensibly due to its environmental credentials, they may feel they have a licence to no longer walk and cycle for certain trips, thus reducing their pro-environmental effort further. Therefore, moral licensing has the potential to increase the size of the observed behavioural rebound effect in the presence of an endogenous increase in energy efficiency.

I develop a novel laboratory experiment to investigate the behavioural rebound effect and moral licensing. The experiment can cleanly isolate pro-environmental behaviours without the many confounds potentially present in the field, such as other motivations to improve energy efficiency or reduce energy usage like saving money. Subjects must decide how to allocate their effort, in a real effort task, between earning money for themselves and avoiding damages to a tree planting charity. I find pro-environmental effort does change with pro-environmental incentives and thus there is a behavioural rebound effect. I also find evidence for moral licensing, particularly for individuals with a stronger pro-environmental orientation of their attitudes and beliefs. Finally, the main driver of pro-environmental effort is beliefs about social norms.

There is a significant literature on pro-environmental behaviours, and how they are driven by preferences and social norms (eg. Costa and Kahn, 2013; Croson and Treich, 2014; Sturm and Weimann, 2006). Allcott and Mullainathan (2010) point to the power of non-price, behavioural interventions in decreasing energy use, compared with improvements in energy efficiency. This paper adds an important new contribution to the empirical literature by looking at resource conservation from the opposite direction, namely the behavioural implications of technology change. To further contribute to this literature, I also measure drivers of underlying willingness to sacrifice for the environment, including pro-environmental orientation of values and beliefs about social norms. Additionally, the experimental design itself is an innovation; I am not aware of any similar laboratory experiments that measure responses to a consumption externality, which involves real world environmental damages.

In the next section I review some background to this study. Section 4.3 follows with an outline of the method, starting with a definition of the behavioural rebound effect in relation to the canonical model of the rebound effect, and concluding by describing the experiment and the hypotheses. In Section 4.4 I present the results, followed by discussion and conclusion sections.

4.2 Background

The existing literature on the rebound effect has identified three levels at which the rebound effect operates - the direct rebound effect, the indirect rebound effect and macroeconomic rebound effects. The direct rebound effect relates to the specific good for which there is an energy efficiency improvement. The direct rebound effect can be defined as the efficiency elasticity of an energy service (Sorrell and Dimitropoulos, 2008). Using the car example, this is the percentage change in kilometres driven divided by the percentage change in energy efficiency. The indirect rebound effect relates to other goods. It is the increase in energy usage from an increase in consumption of other goods after an increase in energy efficiency in one good, which can be modelled as the balancing of income and substitution effects within a consumption bundle (Ghosh and Blackhurst, 2014). Finally, macroeconomic rebounds are due to a reduction of market prices for energy in general equilibrium due to lowered demand after increases in the average level of energy efficiency across the economy. This reduction in market price offsets energy savings as consumption of the energy good is encouraged by the reduction in price (Gillingham et al., 2016). While these latter two types of rebound effects are important for the overall picture, this paper is focused at the level of the direct rebound effect.

I measure just the direct behavioural rebound effect as this type of rebound effect is extremely difficult to measure in the field. Focusing on just the behavioural rebound effect removes potential confounds associated with designing an experiment to also measure the direct rebound effect. Furthermore, direct rebound effects have been estimated in the field for a number of energy-consuming goods, particularly transport and heating. While estimates vary, the average estimated size of the direct rebound effect for household energy services, including driving, tends to be in the range of 5 to 40% (De Borger et al., 2016; Gillingham et al., 2016; Sorrell et al., 2009). It is important to note that the macroeconomic rebound effect could be substantial, with recent dynamic modelling showing backfire is a possibility at the macroeconomic level (Chang et al., 2017).

A range of lab and field experiments have shown individuals will undertake actions for the benefit of others and the public good. Theoretically, intrinsic motivation or environmental preferences can explain some pro-environmental behaviours; other motivations include image, identity and expectations about the motivations and behaviours of others (Ariely et al., 2009; Bénabou and Tirole, 2006, 2011; Bowles and Polanía-Reyes, 2012; Brekke et al., 2003; Nyborg et al., 2006). These theories underpin empirical literature on pro-environmental behaviours. This literature includes evidence that many individuals will pay a premium on particular consumer products for their "green" credentials

(Croson and Treich, 2014). There is also work explaining effort put into recycling, water use reduction and energy conservation using environmental preferences and social norms (Abbott et al., 2013; Allcott, 2011; Ayres et al., 2013; Costa and Kahn, 2013; Ferraro and Price, 2013; Halvorsen, 2008). Important for this paper is that while heterogeneous, many individuals do exhibit a willingness to make some personal sacrifice for the environment (Sturm and Weimann, 2006). Additionally, the fact that environmental behaviours are heterogeneous means questions of heterogeneity in pro-environmental attitudes and behaviours can be explored even with the standard student subject pool, which is otherwise largely homogeneous.

Moral licensing has the potential to increase the behavioural rebound effect associated with technological change when that change is endogenous. Since the first study identifying moral licensing (Monin and Miller, 2001), the effect has been found in a number of studies, within and between a range of domains. Blanken et al. (2015) undertake a meta-analysis of 91 studies and find a small to medium effect of moral licensing, in comparison with other effect sizes of behavioural patterns within the field of social psychology. Domains studied include job hiring, racist attitudes, charitable donations and consumer behaviour. Within environmental economics, Tiefenbeck et al. (2013) find a water conservation campaign in an apartment complex that resulted in a 6% reduction of water use saw electricity use increase by 5.6% for the treatment group, compared with the control group. Moral licensing could increase the rebound effect if an individual purchases a particularly durable good such as a car, and use this purchase to psychologically justify driving more.

Laboratory experiments have been successfully utilised as a method for gaining greater insight into real world economic behaviours in a range of contexts, including environmental economics (Friesen and Gangadharan, 2013; Sturm and Weimann, 2006). A strength of the method is the high level of control it accords the researcher in measuring very specific treatment effects, with a high degree of confidence in claims of exogeneity and a minimisation of potential confounds. This trait makes laboratory experiments particularly suited to investigating behavioural responses to real world phenomena or policies that are difficult to isolate in the field. Limitations of the laboratory environment include the behavioural implications of a high level of salience to subjects of the effect of their actions – in this case environmental damages – and an awareness of being observed (Schubert, 2017; Levitt and List, 2007). Understanding the implications of these limitations has helped guide the experimental design and interpretation of results presented here.

In the case of the rebound effect, behavioural responses to technological change are particularly tricky to identify in the field. Investment by households in durable goods is an endogenous decision, including the choice of level of energy efficiency of a vehicle or appliance (De Borger et al., 2016). Secondary field data has been important in measuring the rebound effect and is indeed the primary means by which the rebound effect is measured. However, for the reasons just mentioned, this is not a straightforward task, meaning there is considerable variance of estimates of the rebound effect in the literature and some methodological debate (Gillingham et al., 2016; Hunt and Ryan, 2014; Sorrell et al., 2009).

Beyond the endeavour of measuring the rebound effect is testing the theory underpinning the hypothesised drivers of the rebound effect. In this case, endogenous investments prove even more problematic to investigating the importance of specific drivers, such as underlying environmental and social preferences and other behavioural phenomena. This is because investment in energy efficiency is likely to be highly correlated with environmental preferences and beliefs about social norms. Research in the lab is a low cost means by which to investigate particular treatment effects, such as response of pro-environmental effort to change in energy efficiency, while ensuring highly credible exogeneity. A carefully considered field experiment into the rebound effect may be highly valuable in this regard too, but a laboratory experiment will increase the evidence base and potentially inform the design for more high cost field work. Thus, a laboratory experiment is highly suited to the research aims of this paper.

4.3 Method

4.3.1 Defining the behavioural rebound effect

I divide this part of the method section into two subsections. First, I discuss the basic definition of the direct rebound effect, given by Sorrell and Dimitropoulos (2008). In the second part I extend the model to include pro-environmental effort to define a behavioural rebound effect and show that it could form an important part of the rebound effect in energy usage.

The basic model of the direct rebound effect

The starting point for defining the rebound effect is to formalise an energy service, ES, as $ES = es[S, A]$. S is useful work (in the physics use of the term, such as kilometres travelled) and A is other attributes of the service (for example comfort). In the basic model, useful work is produced from energy through the following relation:

$$S = \epsilon E. \tag{4.1}$$

The term ϵ is energy efficiency; effectively it is an output-input ratio, which is a function of capital. E is energy, provided by inputs such as petrol or electricity.[1]

An individual decides on the amount of S to consume, given their preferences, budget constraint and the total cost of consuming S. Let S^* be the optimal level of S chosen by the individual. To illustrate further how S^* is chosen, let P_S be the price of the energy component of providing S, which is one component of the total cost of consuming S. Other components of total cost include maintenance of capital and time costs, and are held constant for the purpose of this analysis. The price per unit of energy is given by P_E, and is also held constant. Thus, the energy cost of S is given by:

[1]More generally, E could be any resource for which its use is associated with an environmental externality, such as water. However, I keep with the rebound effect literature by calling this resource energy.

$$C_S = P_S S \tag{4.2a}$$

$$= P_S \epsilon E \tag{4.2b}$$

$$= P_E E, \tag{4.2c}$$

where C_S is the energy cost of S, and thus P_E is the price of energy. Therefore, as shown above, $P_S = P_E/\epsilon$. This relationship between the change in the energy cost of S and a change in energy efficiency, ϵ, is what the rebound effect hinges on. While a number of variables, including P_E, will affect the optimal choice of S, S^*, all variables are held constant except ϵ in this analysis. Therefore, I focus on the effect of ϵ on S^* through the function $S^*(\epsilon)$.

With no change in S^* after an increase in ϵ, there is no rebound effect; energy use decreases in proportion to any increase in energy efficiency. However, S^* may increase after an increase in ϵ, holding the price of energy, P_E, constant. An increase in ϵ reduces P_S, and thus can increase S^* through positive income and substitution effects.[2] In this case, there is a positive rebound effect. Furthermore, this line of reasoning shows that S^* can be thought of as a function of ϵ, through the effect of ϵ on P_S.

Rearranging equation (4.1) and taking the derivative of E with respect to ϵ, we get the change in energy use in response to a change in energy efficiency:

$$\frac{\partial E}{\partial \epsilon} = -\frac{S^*(\epsilon)}{\epsilon^2} + \frac{1}{\epsilon}\frac{\partial S^*(\epsilon)}{\partial \epsilon}. \tag{4.3}$$

Assuming an increase in ϵ, the first right hand side term of this equation is the direct change in energy use due to a change in energy efficiency, assuming no change in S^*. This term can thus be interpreted as the change in energy use due to simple engineering calculations. The second term on the right hand side of the equation is the increase in energy use due to an increase in S^* after an improvement in energy efficiency. Thus, this

[2]See Chan and Gillingham (2015) for a full derivation of the rebound effect using utility theory. Consistent with the literature on the rebound effect in general, they do not include pro-environmental preferences or social norms.

second term is the increase in energy use due to the direct rebound effect. The size of this term is determined by the size of the income and substitution effects - how much the individual puts their saved income into consuming more S versus other goods. If there is no rebound effect, there are no income and substitution effects, S^* is no longer a function of ϵ and this last term in equation (4.3) falls away.

The direct rebound effect is specifically defined as the proportional increase in useful work from the energy service consumed relative to the proportional increase in energy efficiency. This is equivalent to the efficiency elasticity of demand for useful work:

$$\eta_\epsilon(S) = \frac{\partial S^*(\epsilon)}{\partial \epsilon} \frac{\epsilon}{S^*(\epsilon)}. \tag{4.4}$$

In the absence of a direct rebound effect, all improvements in energy efficiency lead to a 1 for 1 reduction in energy use. In this case, $\eta_\epsilon(S) = 0$. With a positive direct rebound effect, $\eta_\epsilon(S) > 0$. Backfire occurs when the direct rebound effect is so great that energy usage actually increases after an improvement in energy efficiency, in which case $\eta_\epsilon(S) > 1$.

The behavioural rebound effect

I now extend the basic definitions to include pro-environmental effort, in order to define the behavioural rebound effect. Pro-environmental effort is undertaken to conserve energy for environmental reasons, for example riding a bicycle to avoid consuming petrol by driving. Pro-environmental effort is positive when individuals are sufficiently motivated by their pro-environmental preferences or preferences to conform with social norms, given the costs (monetary or otherwise) of undertaking such effort. An important relation underpinning this extended model is the negative association between the efficiency of pro-environmental effort and energy efficiency, ϵ. When the energy efficiency of a car improves, the amount of energy saved per kilometre by riding a bicycle falls. Many other pro-environmental behaviours in this example also follow this logic - keeping tyres inflated or having a light foot on the accelerator also save less petrol per kilometre with an efficient car compared with an inefficient car.

To consider the extended model more formally, let M be pro-environmental effort. I define M such that it only incorporates effort expended for environmental reasons - either due to pro-environmental preferences or social norms. The term M does not include ostensibly pro-environmental effort, such as riding a bike, where that effort is done to advance other objectives, such as to save money, for enjoyment or to get fit.

Let the energy conserved by pro-environmental effort, E_M, be given by:

$$E_M = \phi M. \tag{4.5}$$

The term ϕ is the efficiency of pro-environmental effort in reducing energy usage, effectively an output-input ratio of energy savings from pro-environmental effort. In this extended model, the energy used by consuming useful work S, previously defined by rearranging equation (4.1), is given by:

$$\begin{aligned} E =& \frac{S}{\epsilon} - E_M && \text{(4.6a)} \\ =& \frac{S}{\epsilon} - \phi M. && \text{(4.6b)} \end{aligned}$$

Hence, pro-environmental effort is a substitute for energy, E, which is defined as an environmentally damaging energy source, like petrol. Useful work consumed, S^*, is assumed to be a function only of ϵ, and is not affected by pro-environmental preferences or social norms for pro-environmental effort. Therefore, S^* in this model can be interpreted as useful work consumed in absence of pro-environmental preferences and social norms. Optimal level of pro-environmental effort, M^*, is a function of ϕ as the level of pro-environmental effort depends on the efficiency of pro-environmental effort, given pro-environmental preferences and social norms, and the private costs incurred from undertaking pro-environmental effort. Preferences and effort costs are held constant. It is thus assumed that pro-environmental effort, M, is the channel through which individuals reduce their damage to the environment.

As noted at the start of this section, ϕ is a function of ϵ such that:

$$\frac{\partial \phi(\epsilon)}{\partial \epsilon} < 0. \tag{4.7}$$

Thus, an improvement in energy efficiency reduces the benefits from undertaking a pro-environmental behaviour.

I can now derive the equation in the extended model that is equivalent to equation (4.3) in the basic model:

$$\frac{\partial E}{\partial \epsilon} = -\frac{S^*(\epsilon)}{\epsilon^2} + \frac{1}{\epsilon}\frac{\partial S^*(\epsilon)}{\partial \epsilon} - \frac{\partial \phi(\epsilon)}{\partial \epsilon} M^*(\phi(\epsilon)) - \phi(\epsilon)\frac{\partial M^*(\phi(\epsilon))}{\partial \phi(\epsilon)}\frac{\partial \phi(\epsilon)}{\partial \epsilon}. \tag{4.8}$$

Note that S^* and M^* are jointly chosen to maximise the individual's utility, hence I consider their simultaneous effect on change in E. The first two terms on the right hand side of the equation are unchanged from the base model, as shown in equation (4.3), however their interpretation changes slightly. The first term on the right-hand side is now just one part of the engineering calculation. The engineering calculation must also include the third term on the right hand side of the equation. This term is the change in energy conserved given a change in energy efficiency, but no change in pro-environmental effort.

The second term on the right hand side of equation (4.8) is the resulting change in energy use due to an increase in consumption of useful work from the energy service; termed the direct rebound effect, as before. This term only incorporates the change in consumption of useful work from the energy service due to private income and substitution effects and does not include pro-environmental preferences or preferences to avoid deviations from social norms. The last term on the right hand side gives the change in M^* caused by an increase in ϵ, which is a result of what I call the behavioural rebound effect. If $\frac{\partial M^*(\phi(\epsilon))}{\partial \phi(\epsilon)} > 0$, then this final term in equation (4.8) is also positive, hence there is a positive behavioural rebound effect. That is, the change in pro-environmental behaviours leads to less energy savings from an improvement in energy efficiency than predicted solely by the engineering calculations. Therefore, this extended model separates out the direct rebound effect, as influenced by private income and substitution effects, and the behavioural rebound effect, which is influenced by the effect a change in energy efficiency has on incentives for pro-environmental effort. The combination of these two rebound effects determine the overall rebound effect as it pertains to energy use, E.

It is important to emphasise that this model hinges on the definition of M as pure pro-environmental effort. In my experiment I can measure pro-environmental effort directly, hence it is useful to separate the direct rebound effect from the behavioural rebound effect. However, in the field it would be difficult to measure M specifically. For example, in the base model, some pro-environmental effort would be captured by a higher ϵ. Using the transport example, this could be ensuring tyres are fully inflated or using a light foot on the accelerator pedal. Other pro-environmental effort would be captured through a lower S^*, such as reducing distance driven, through cycling or substituting driving with other activities; again, purely for positive environmental outcomes. Therefore, this extended model is intended to complement the existing literature on the rebound effect by providing a formulation that allows for behavioural rebounds to be explicitly included and hence test their importance for rebounds in energy usage.

Thus, the extended model defines a behavioural rebound effect, equivalent to the negative of the energy efficiency elasticity of pro-environmental effort:

$$-\eta_\epsilon(M) = -\frac{\partial M^*(\phi(\epsilon))}{\partial \epsilon}\frac{\epsilon}{M^*(\phi(\epsilon))} \tag{4.9a}$$

$$= -\frac{\partial M^*(\phi(\epsilon))}{\partial \phi(\epsilon)}\frac{\partial \phi(\epsilon)}{\partial \epsilon}\frac{\epsilon}{M^*(\phi(\epsilon))}. \tag{4.9b}$$

Hence, there is no behavioural rebound effect when $-\eta_\epsilon(M) = 0$, a positive behavioural rebound effect when $-\eta_\epsilon(M) > 0$, and backfire when $-\eta_\epsilon(M) > 1$. This extended model now implies the rebound effect in energy use is the sum of two separate rebound effects.

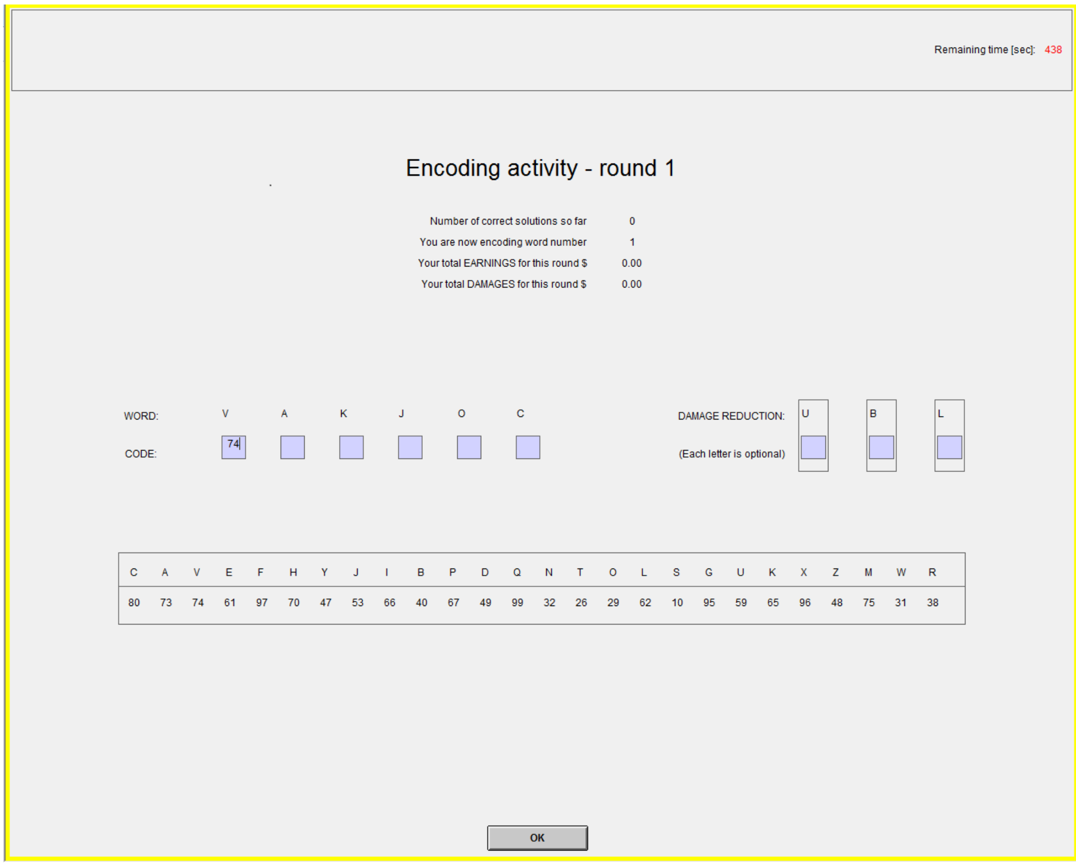
Moral licensing has the effect of increasing the size of the behavioural rebound effect relative to when no moral licensing has occurred. After an individual makes a moral choice, moral licensing is revealed as a subsequent immoral action or a decrease in the level of moral effort the individual otherwise would have made. Thus, if there is a larger behavioural rebound effect after an endogenous increase in ϵ compared with the same change in ϵ imposed exogenously, then moral licensing has occurred. This comparison must be done with equal costs for the change in ϵ to ensure it is not the cost of the choice that is driving the reduction in pro-environmental effort.

The main aim of this experiment is to estimate the behavioural rebound effect, $-\eta_\epsilon(M)$. Through estimating the behavioural rebound effect I can test whether pro-environmental effort, M, is an important part of the overall rebound effect in energy use, E. The experimental design allows me to estimate the behavioural rebound effect without confounding it with the direct rebound effect, $\eta_\epsilon(S)$. Thus, my experimental design is aimed at measuring just the behavioural rebound effect; it is beyond the scope of this paper to measure the full rebound effect in E, in a laboratory setting. By measuring the behavioural rebound effect I can compare its magnitude to the direct rebound effect of energy use as measured in prior research in the field. Given the behavioural rebound effect can also be decomposed into income and substitution effects as they relate to trading off private consumption and reducing environmental damages, I additionally measure just the income effect component of the behavioural rebound effect. Another important component of the behavioural rebound effect is $\eta_\phi(M)$, which is the efficiency elasticity of pro-environmental effort. This elasticity has a direct impact on the size of the behavioural rebound effect, as follows from equation (4.9). Hence I measure $\eta_\phi(M)$ directly, without an associated change in ϵ. Finally, I test whether there are moral licensing effects, which are shown if $-\eta_\epsilon(M)$ with an endogenous increase in ϵ is greater than $-\eta_\epsilon(M)$ with an exogenous increase in ϵ.

4.3.2 Experimental design

The basic design of the experiment allows the estimation of how subjects trade off between their consumption (monetary earnings) and environmental damage (reduction in a donation to a tree planting charity). By varying damages between rounds (within subjects), I can estimate the size of the subjects' behavioural rebound effect. By varying the treatments shown to subjects, the experimental design also allows testing between

Figure 4.1: Experimental screen of the main activity.



subject hypotheses, such as that there is moral licensing. To link the experiment with the model, monetary earnings before any sacrifice for the environment can be thought of as S, environmental damages can be thought of as E, with ϵ determining the level of damages associated with S. Damages can be reduced through pro-environmental effort, M at a relative cost of ϕ.

The experimental activity is based on a word decoding effort task, similar to Erkal et al. (2011) and Benndorf et al. (2014). At the start of each eight minute round, subjects are presented with a screen as shown in Figure 4.1. Subjects must correctly enter the two digit codes for each of the random letters for the six letter "word" they are given. The codes are provided in a scrambled alphabet across the bottom of the screen. This word is displayed in the centre left of the screen. Once a subject has correctly completed the word, she can click the OK button and earn the payment for that word - which is 60c for most treatments. Thus, for the subject to maximise her earnings for the round, she must try to complete as many six letter words as possible within the eight minute time limit.

Each completed word reduces a charity payment for that round. The charity is a local tree planting charity, and subjects know that every $2 donated to the charity leads to one seedling being planted. In the high damage treatment, each word completed reduces the charity payment by 54c. However, in the centre right of the screen, subjects can lower the damages to the charity for that word by filling in additional letters. It is made clear to the subjects that these additional letters are optional, both on screen and in the instructions. One additional letter will lower the damages for that word by one third, two by two thirds and all three additional letters will lower the damages to the charity to nothing. As filling in the additional letters takes extra time, subjects must trade off how much damage they are willing to do to the charity payment (the environment) with their private earnings in each round. Cumulative earnings and damages for the current round are displayed in the top centre of the screen. The full instructions are provided in the Appendix.

A real effort task was chosen for the experimental activity given pro-environmental effort in the field requires both real effort in a task, as well a sacrifice of private consumption. The experimental design ensure subjects face both these costs in the laboratory too.[3] Additionally, consumption requires income, garnered through effort, and consumption of useful work of an energy service, S, may also require a labour input such as driving. All of these types of effort are accounted for in the experimental design.

The word decoding task is a modified version of Erkal et al. (2011), such that each word is composed of six letters, plus two or three optional extra letters to reduce environmental damages. Furthermore, the order of the alphabet is scrambled for each word, as suggested by Benndorf et al. (2014). This scrambling is done to minimise any learning effects between rounds and was successful in this case as no learning effect was observed (see Section 4.4.1). Finally, none of the eight or nine letters given to subjects to decode for each word were repeated within that word. In the piloting stage of the experiment it was observed that subjects were more likely to complete the optional extra letters if they were repeats of letters already given for that particular word.

[3]The experimental design means there is a clear opportunity cost of private consumption to completing the optional extra letters. This design helps to ensure the measurement variable of interest, pro-environmental effort, is responsive enough to changes in incentives given the sample size, which is not always the case when using real effort tasks (Araujo et al., 2016; Erkal et al., 2017).

Each session consisted of 24 subjects. The donation to the charity was made at a session level. The initial donation for any given round was set at \$336; this fact was communicated to subjects in the instructions. This amount meant that there was \$14 donated per subject, which was high enough to ensure the session level charity donation was not depleted to \$0 for a given round, therefore ensuring marginal damages were never 0. Each subject completed a practice round of eight minutes, plus three rounds of eight minutes each. One of the three rounds was randomly chosen and paid out.

There were five treatments given to subjects. Each subject was given one treatment per round, thus each subject saw three treatments. The treatments are shown in Table 4.1, including payoff per word, damages per word, number of optional extra letters per word and thus implied energy efficiency and efficiency of pro-environmental effort. The equivalent term in the theoretical model is given. Payment per word is equivalent to one unit of S in the theoretical model, when no optional additional letters are completed. As energy consumption is associated with damages, damages per word is equivalent to E consumed per unit of S, expressed in terms of environmental damages. The optional extra letters per word provide a maximum level of pro-environmental effort, M, that is possible per unit of S. Energy efficiency is calculated according to $\epsilon = S/E$, when no optional additional letters are completed, hence it can be calculated by dividing payoff per word by maximum damages per word. Finally, efficiency of pro-environmental effort, ϕ, is the ratio of reduction in damages to sacrifice of consumption given $\phi = E_M/M$, as per equation (4.5).[4] Private earnings for a round, Y, for subject i is determined by $Y_i = S_i - M_i$, where S_i is total round earnings absent pro-environmental effort and M_i is earnings sacrificed for the environment.

Running through the treatment values shown in Table 4.1, high damage, low damage,

[4] I provide an example of how ϕ is calculated using the high damage treatment (see Table 4.1). To determine E_M/M both E_M and M must be put into an equivalent unit, money. One unit of M is one extra letter, thus in monetary terms it is equivalent to sacrificing 1/6 of the earnings per word, or \$0.10 in the high damage treatment. The damage reduction from one unit of M, or E_M, is the sum of two values. The first part of E_M is 1/3 of the damages per word, as explained above, or \$0.18 in the case of the high damage treatment. However, it also reduces the amount of words the subject can complete within the eight minute time limit by 1/6. This gives an additional damage reduction of 1/6 of the damages caused by a word with no pro-environmental effort. Hence, for the high damage case, $E_M = \$0.18 + \$0.09 = \$0.27$. Thus, $\phi = 2.7$ for the high damage treatment.

Table 4.1: Treatment parameters.

Experimental parameter	Payoff/word	Damage/word	Optional letters/word	Energy efficiency	Efficiency of pro-env. effort
Theoretical interpretation	Unit of S	E consumed per unit of S	Max M per unit of S	ϵ	ϕ
Treatment					
High damage	\$0.60	\$0.54	3	1.1	2.7
Low damage	\$0.60	\$0.36	3	1.7	1.8
Choice	\$0.60	\$0.54 or \$0.36	3	1.1 or 1.7	2.7 or 1.8
Low effort	\$0.60	\$0.54	2	1.1	3.6
High income	\$0.80	\$0.72	3	1.1	2.7

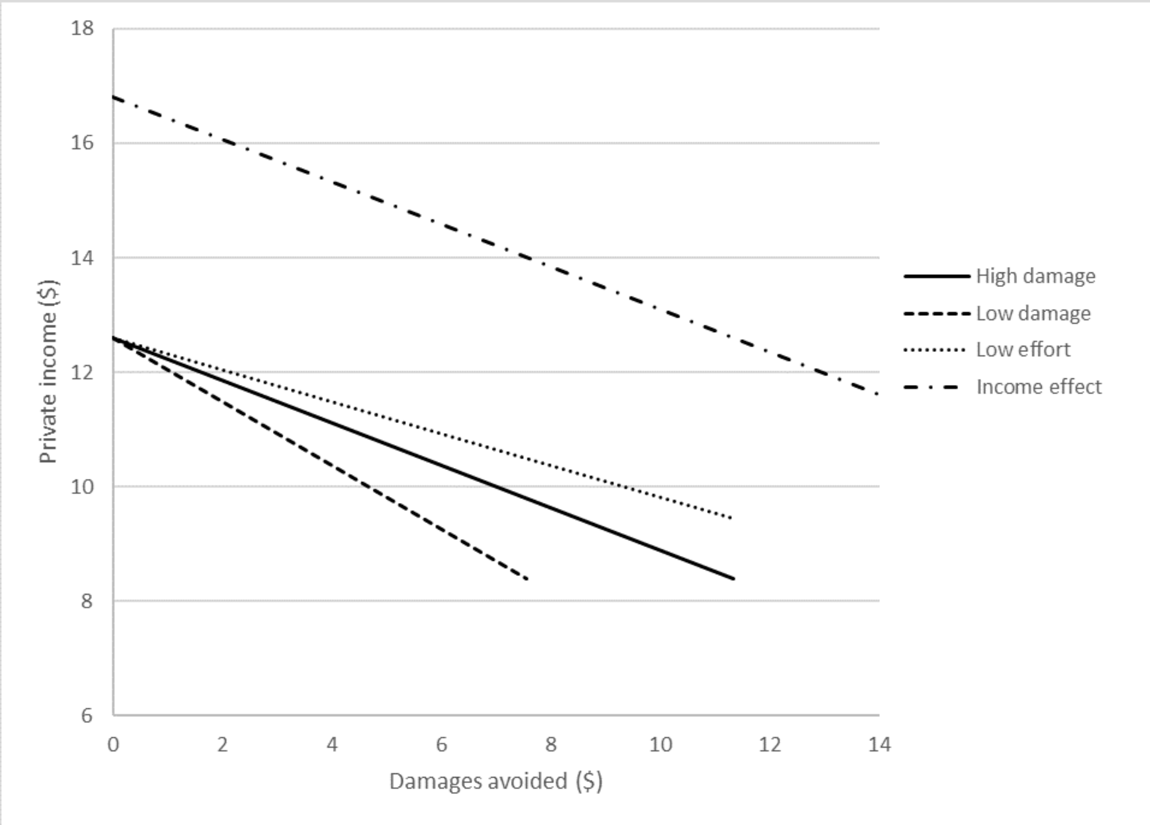
choice and low effort treatments all pay \$0.60 per word, but vary by damages and number of optional extra letters per word. High damage treatment has damages of \$0.54 per word, whereas low damage has damages of \$0.36 per word. Choice tests for moral licensing - at the start of the round, subjects are given the costless choice of causing either \$0.54 or \$0.36 of damages per word.[5] Low effort tests what happens when ϕ is increased without an increase in ϵ. This increase in ϕ is achieved by lowering the number of optional extra letters from three to two, where one extra letter completed lowers the damages by half, and two extra letters lowers the damages to 0. Thus, damages per word are the same as high damages, so ϵ is unchanged, whereas ϕ increases. Finally, the high income treatment provides a test of pure income effects - payoff per word and damages per word are both increased by one third relative to high damage.

By design $\phi > 1$ for each treatment to ensure total welfare within the experimental session (subject payoffs plus donation to the charity) is highest when subjects always complete all optional extra letters. Thus, ϕ is akin to the multiplier used in standard experimental games, such as public good and trust games, where donations to a public good or to other players are increased in value by the experimenter (Berg et al., 1995; Sturm and Weimann, 2006). Also note that $\phi = 3/\epsilon$ except for the low effort treatment where $\phi = 4/\epsilon$.[6] Hence, the assumption given in equation (4.7) holds.

[5] The choice is costless in order to ensure there are no income effects confounding the difference between pro-environmental effort in the low damage treatment and those who chose low damages, as noted in Section 4.3.1.

[6] These relationships between ϕ and ϵ follow from equation (4.6b). To solve for ϕ as a function of ϵ, consider the completion of one word with all the optional extra letters completed. Keeping with consistent units, this sets $M = M_{max}$ per word (either \$0.20 or \$0.30), $E = 0$ and $S = \$0.90$, as S is the monetary

Figure 4.2: Budget constraints by treatment, faced by a subject who can complete 126 letters in eight minutes.



The tradeoffs faced by subjects in each treatment can be represented as a budget constraint, as shown in Figure 4.2. The example shown represents a subject who is able to complete 126 letters within the eight minute time period, and graphs the various allocations of letters between reducing damages to the environment (E_{Mi} on the x axis) and private income (Y_i on the y axis).[7]

There are five treatment groups, grouped by the treatments and the order of treatments the groups received. These groups are shown in Table 4.2, along with the number of subjects in each group. These treatment groups allow for the testing of between subject hypotheses using a differences-in-differences approach. Specifically, comparing the difference in pro-environmental effort between treatment groups A and B with C for rounds 1 and 2 allows for the testing of order effects to ensure they do not play a role in driving

value of the letters completed in absence of pro-environmental effort. Thus, ϕ can be solved for as a function of ϵ.

[7] The slope of the lines is given by $1/\phi$. The slope is calculated by taking subject earnings for a round, $Y_i = S_i - M_i$. Earnings sacrificed for pro-environmental effort, M_i, can be substituted for from equation (4.5), giving the equation for the relation between earnings and damages avoided, $Y_i = S_i - \frac{1}{\phi} E_{Mi}$.

Table 4.2: Treatment groups by treatment order plus number of subjects in each group.

Treatment group	Round 1	Round 2	Round 3	Number of subjects
A	High damage	Low damage	Low effort	47
B	High damage	Low damage	High income	24
C	Low damage	High damage	Low effort	48
D	High damage	Choice	Low effort	48
E	High damage	Choice	High income	48
Total subjects				215

the overall results. Comparing the difference in pro-environmental effort between rounds 1 and 2 between treatment groups A and B and treatment groups D and E allows for the testing of moral licensing.[8]

Experimental procedures

The 9 experimental sessions of 24 subjects each were conducted at the Monash Laboratory for Experimental Economics (MonLEE) at Monash University in Melbourne, Australia.[9] Current students of Monash University, registered in the MonLEE subject pool, were invited to attend a maximum of one session each. Sessions were conducted in June 2016 and May 2017. Student were invited to participate using ORSEE subject management system for the 2016 sessions (Greiner, 2015) and SONA for the 2017 sessions.[10] The study was named "A study of behaviours" so that the recruitment process was not biased towards students with an interest in environmental issues.

At the beginning of the session time, after being checked off from the attendance role, subjects were allowed to sit at any available computer. Computers at MonLEE are set up with screens so that subjects cannot see neighbouring screens and no communication between subjects was permitted. Subjects read a generic explanatory statement and signed a consent form. Once the consent forms were completed and signed, the

[8] Only one session of 24 subjects was required for treatment group B as this group is not used to test any hypotheses on its own. Treatment group B boosts the numbers for the high damage to low damage combination from round 1 to 2 of treatment group A, and the high damage to high income combination from round 1 to round 3 combination of treatment group E.

[9] One session for treatment group A had only 23 subjects attend. Due to this lower number of subjects, the charitable donation for the rounds for that session was lowered by $14 and this was explained to subjects at the start of the session.

[10] The transition from ORSEE to SONA was managed such that no subject could participate in the study twice. See https://www.sona-systems.com/default.aspx for information about SONA (accessed 29 April 2017).

overview instructions were read, followed by the activity instructions (see Appendix). Next, subjects undertook a simple and incentivised quiz to ensure they understood the instructions. There were four questions, some with multiple parts; a fully correct answer for one question earned subjects 25c. Subjects were informed by the software immediately after submitting their answer whether or not they were correct – if incorrect, the correct answer was given and explained.

After all subjects finished the quiz, the activity instructions were given and the activity commenced. It was explained in the activity instructions that there were to be three rounds of eight minutes each, with one round being randomly chosen by the computer to be paid. The earnings and damages per word for each round were read aloud and displayed on the screen before each round to establish common knowledge that every subject had the same incentives for the round. After the three rounds, a survey was given to subjects – the variables used from the survey in the analysis are described in Section 4.4.1.

All activities and the survey were conducted using the z-Tree program (Fischbacher, 2007). After the survey, the experimenter announced the round that would be paid, including the total session-level payment to the charity. It was explained at the start of the activities that the charity payment would not be known until this point, and that proof of the donation would be provided via email in the days after the experiment had finished. Finally, each subject was paid in private in Australian dollars. Each session lasted roughly one hour.

4.3.3 Hypotheses

Within subject hypotheses

The hypotheses are described here in relation to the theoretical model of Section 4.3.1. They are separated into within and between subject hypotheses. The first within subject hypothesis is as follows:

H1 The behavioural rebound effect is positive.

The behavioural rebound effect is equivalent to the negative of the energy efficiency elasticity of pro-environmental effort, $-\eta_\epsilon(M)$. That this value is positive follows from its definition in equation (4.9), and the assumption given in equation (4.7). Specifically, an increase in energy efficiency, ϵ, has a negative impact on the efficiency of pro-environmental effort, ϕ, by assumption and by experimental design. A decrease in ϕ reduces the benefit/cost ratio for pro-environmental effort, which I hypothesise will lead to a decrease in pro-environmental effort, M. This decrease in pro-environmental effort is consistent with assuming pro-environmental effort is undertaken both for pro-environmental preferences and beliefs about pro-social norms. Thus, hypothesis H1 is that $-\eta_\epsilon(M) > 0$.

Hypothesis H2 is related to the difference in the treatments where ϕ is varied but ϵ is the same (high damage compared with the low effort treatments). Thus, looking just at the efficiency elasticity of pro-environmental effort, $\eta_\phi(M)$, and consistent with hypothesis H1, the second hypothesis is:

H2 The efficiency elasticity of pro-environmental effort, with no change in energy efficiency, is positive.

Finally, hypothesis H3 is related to the change in pro-environmental effort between the high damage and high income treatments. As the behavioural rebound effect is an elasticity, it is composed of an income effect and a substitution effect, which could operate in the same or opposite direction. Hence, I estimate the income elasticity of pro-environmental effort, call it $\eta_Y(M)$, to determine the magnitude and direction of the income effect for $\eta_\phi(M)$. Thus, this hypothesis allows me to approximate the size of the substitution effect. The difference in pro-environmental effort between the high income treatment and the high damage treatment tests the direction of the income effect in this context. The Environmental Kuznet's Curve hypothesis suggests that pro-environmental preferences rise with income as environmental quality is a luxury good (Dinda, 2004). I thus hypothesise that:

H3 There is a positive income elasticity of pro-environmental effort.

Between subject hypotheses

The first between subject hypothesis relates to the level of pro-environmental effort measured within a given treatment round. It is that:

H4 Pro-environmental effort can be partially explained by demographics, environmental values and beliefs about social norms.

Given the literature outlined in the introduction and background sections, in Section 4.3.1 I argue that level of pro-environmental effort is driven by pro-environmental preferences and aversion to deviating from social norms. I measure beliefs about social norms directly and thus hypothesise that higher levels of beliefs about social norms of pro-environmental effort will drive higher levels of pro-environmental effort. Pro-environmental preferences will likely be formed through a combination of life experience and pro-environmental beliefs and values. There will be little variation in demographics in the data, given the relatively homogeneous subject pool, but measures of pro-environmental orientation along with observed pro-environmental behaviours are still likely to be highly heterogeneous, given findings of other studies (Hawcroft and Milfont, 2010; Sturm and Weimann, 2006). I describe the relevant variables used for this hypothesis in Section 4.4.1.

The next between subject hypothesis is that:

H5 There is a moral licensing effect. Specifically, the drop in pro-environmental effort will be less when moving from the high to low damage treatments compared with the high to low damage choice treatments.

This hypothesis follows from the moral licensing literature, as previously described. It is tested by comparing the treatment groups who were given the low damage treatment exogenously to the treatment groups who were given a costless choice between high and low damages per word. The choice is costless, meaning that the only difference between the two treatments is that the choice itself is the only difference between the two treatments. There are no differences in earnings for that round between the subject groups. Hence, subjects may give themselves a moral licence to put in less effort after choosing low damages compared with when they have low damages imposed exogenously. I test this hypothesis using a differences-in-differences approach, as stated in the hypothesis itself.

Finally, I look at whether there is heterogeneity in the moral licensing effect due to pro-environmental orientation of their attitudes, values and beliefs. Moral licensing occurs when an individual undertakes a moral action - in this case, choosing low damages over high damages for a round - and then give themselves a psychological licence to undertake less moral behaviours after that point than they otherwise would. This effect thus hinges on the individual seeing the action they have undertaken as moral in the first place. Thus, I hypothesise that subjects with a higher pro-environmental orientation will see choosing low damages as more of a moral choice than those with a lower pro-environmental orientation, where pro-environmental orientation is a mix of values and beliefs about the environment. Hence, the final hypothesis is:

H5a The moral licensing effect is larger for those with a higher pro-environmental orientation.

4.4 Results

In the first part of this section I present the summary statistics while clearly defining the relevant variables. Next, I present the econometric analysis of the results. I finish the section by outlining each main result as it pertains to the relevant hypothesis.

4.4.1 Summary statistics

The outcome variable of interest in this experiment is proportion of pro-environmental effort. This variable is calculated as the proportion of the optional extra pro-environmental letters completed out of the total possible, for each individual in each round.[11] As this measure is robust to number of letters completed in a round, it is suitable to use to compare both within individuals (it allows for variation of letters completed within rounds, for example due to an error in one of the words) and between individuals (it allows for difference in overall effort and/or skill). As it is a proportion, it is a continuous variable on the unit interval. This variable is used as the dependent variable throughout the results section.

Proportion of pro-environmental effort by treatment is summarised in the top half of Table 4.3a. The treatment with the lowest effort is choice - chose low, which is the treatment where subjects could choose between high and low damages and includes just those who chose low damages (85 out of the 96 subject given this treatment). Proportion of pro-environmental effort in this treatment is 0.23, meaning less than 1 of the 3 optional damage reducing letters were completed per word. Next lowest treatment by pro-environmental effort is high income, followed by low damage, high damage, low effort and finally choice - chose high.

It is useful to note at this stage that no learning effect is observed; the mean total number of letters completed per round is almost identical over each round. The mean letters completed for all treatment groups in order of round are 125.7, 126.0 and 126.3, with no statistical difference detected ($0.72 > p > 0.88$, depending on the rounds compared). Mean letters completed by treatment group are similarly stable. Thus, the randomised

[11]The variable is calculated so that it takes into account that individuals who complete more optional extra letters will complete the roughly same number of letters in a round, but will complete fewer words, as the optional extra letters make the words longer. Consider an individual who completes 126 letters in a high damage treatment round. If they complete all the optional extra letters, they will complete 14 words ((126/9), which is $14 \times 3 = 42$ extra letters. Thus, the maximum number of extra letters they could complete at a rate of 126 letters per round is 42. Hence, if they complete 1 optional extra letter per word, they complete $126/7 = 18$ words, thus 18 optional extra letters and $18/42 = 0.43$ optional extra letters out of the total they could complete. This example demonstrates that if an individual completes 1 of 3 optional extra letters for each word they complete, rather than the proportion of optional extra letters being completed being 0.33, it is actually 0.43 as the calculation needs to take into account the total number of words they complete is falling as they complete more extra letters.

Table 4.3: Summary statistics.

(a)

Statistic	N	Mean	St. Dev.	Min	Max
Proportion pro-environmental effort by treatment					
High damage	215	0.32	0.39	0	1
Low damage	119	0.29	0.40	0	1
Choice – chose low	85	0.23	0.35	0	1
Choice – chose high	11	0.42	0.45	0	1
Low effort	143	0.36	0.40	0	1
High income	72	0.25	0.36	0	1
Continuous covariates					
Age	215	21.81	3.77	17	48
Norm belief	215	1.11	0.96	0	3
Letters high damage	215	125.77	21.14	63	168
Environmental behaviours	215	3.64	0.44	2.62	4.93
NEP scale	215	3.72	0.47	2.60	4.87

(b)

Statistic	N	%
Gender		
Female	110	51.2
Male	105	48.8
Subjective personal income		
Low	182	84.7
Medium	33	15.3
High	0	0
Citizenship		
Australian	48	22.3
Not Australian	167	77.7
Environmental organisation		
Not member	182	84.7
Member	33	15.3
Political party		
Not member	206	95.8
Member	9	4.2
Voting preference		
Liberal	36	16.7
Labor	27	12.6
Greens	19	8.8
Other	7	3.3
Unsure	126	58.6
Total	215	100

alphabet design from Benndorf et al. (2014) successfully prevented learning effects from affecting the results.

The other summary statistics in Table 4.3 are primarily the subject responses to the survey given at the end of each experimental session. The bottom half of Table 4.3a gives the summary statistics for the continuous covariates. Most subjects are close to the mean age of 21.8, as expected from a standard student subject pool. The norm belief variable gives the subject response to the question of what they believe to be the average number of optional extra letters in round 1 of other subjects in their session. Subjects could only answer in whole numbers between 0 and 3, and on average guessed the correct number, 1. The letters high damage variable is the number of letters completed by subjects in the high damage treatment. Environmental behaviours is a measure of stated frequency of undertaking pro-environmental behaviours within the last year, between 1 (never) and 5 (always). The measure is produced by averaging the reponse to all the environmental behaviour questions included in the survey, for which a Likert scale was employed. Finally, the New Ecological Paradigm (NEP) scale is a measure of pro-environmental orientation of attitudes and beliefs (see Appendix for questions used for these latter two variables). This is also a variable utilising a Likert scale from 1 to 5, depending on answers to a standard 15 question survey on environmental values and attitudes, where a higher number denotes a stronger pro-environmental orientation (Dunlap et al., 2000). The mean value of 3.7 falls within 0.1 of the mean value recorded for two 15 question NEP surveys undertaken in Australia in roughly the last decade (Hawcroft and Milfont, 2010).

Table 4.3b shows the discrete variables. First, the gender balance is very even, with 51% of subjects being female. Subjective personal income level stated by subjects is mostly low (85%), with the rest being medium. This pattern is not unexpected with a student subject pool. Subjective variables such as level of income often prove to be informative explanatory variables (Bertrand and Mullainathan, 2001). Next, the sample has a large number of subjects who are not Australian citizens (78%), which simply reflects the subject pool at the MonLEE lab. There is no particular reason to believe using a largely non-Australian subject pool would affect the testing of the hypotheses, but with collecting data on citizenship I can control for this variable. The next two variables are subject responses to whether they have ever been a member of a environmental organisation or political party, to indicate political engagement, particularly concerning environmental issues. Not many subjects report being or having been a member of either (15% and 4% respectively). Finally, the voting preference question asked subjects which political party they would give their first preference to if voting on the day of the survey, and regardless of their Australian citizenship status. Of note to the research question is 9% stating they would vote for the Greens Party. A majority stated they were unsure at 59%, which is unsurprising given the large number of non-citizens.

4.4.2 Econometric analysis

Here I outline the main econometric approach and introduce the overall results. In Sections 4.4.3 and 4.4.4 I discuss in detail the results as they pertain to each hypothesis.

The within subject hypotheses depend on the difference in pro-environmental effort between particular treatments; the difference from high to low damage treatments, the difference between high damage to low effort treatments and the difference between high damage and high income treatments. Thus, I test whether the differences in pro-environmental effort between these treatments are statistically significant and in the direction consistent with the first three hypotheses. Table 4.4 tests these differences using the non-parametric Mann-Whitney U test. The first row in Table 4.4 tests the difference between the high damage and low damage treatments, specifically testing whether the proportion of pro-environmental effort in the low damage round minus the high damage round is negative. This test is done for all 119 subjects who received both treatments. The result is negative, with a p-value of 0.001. The second row in Table 4.4 tests the difference between the high damage and low effort treatments, where the low effort treatment has the same damage level as the high damage treatment per word, but only requires 2 optional additional letters to be completed to reduce damages to 0 for each word, rather than 3. The final row tests tests the difference between the low and high income treatments, specifically testing whether pro-environmental effort in the high income round minus the high damage round

Table 4.4: Testing for differences between treatments in the direction relevant to the within subject hypotheses, using the non-parametric paired Wilcoxon signed-rank test.

	Negative	Positive	N
Difference			
High to low damage	0.001***		119
High to low effort		0.185	143
Low to high income		0.738	72

Note: *p<0.1; **p<0.05; ***p<0.01.

is positive. Neither of these two values are found to be statistically positive.[12]

Estimated treatment effects are shown in Table 4.5. The first two columns show model (1), which regresses dummy variables for each treatment round, relative to high damage, and each treatment group, relative to treatment group A, on the proportion of pro-environmental effort. A Tobit model is used given the dependent variable is subject to corner solutions (Wooldridge, 2010). I discuss this choice of model in more detail shortly. The left column for model (1) shows the estimated coefficients, whereas the right column shows the average marginal effects (AME). Average marginal effects are of more interest in this paper as they show the average effects of the treatments for the subject pool on the proportion of pro-environmental effort.[13] Each model in Table 4.5 uses all three observations from all 215 subjects, thus standard errors are clustered at the subject level.

Focusing on the low damage and low effort coefficients in model (1), Table 4.5, the former is negative and statistically significant and the latter is positive but not statistically significant, consistent with the results in Table 4.4. The treatment group controls are important to include to remove any differences between the average effort levels of

[12]It could be argued that total effort (total number of letters completed within a round) could also be increased from low to high income treatments, given the higher reward per letter in the high income treatment. I test for whether there is an increase in total effort between these treatments using a one-sided, paired Wilcoxon signed-rank test, and find it is not statistically significant ($p = 0.161$), therefore it does not affect the analysis in any significant way.

[13]The Tobit coefficients can be interpreted as the estimated marginal effect of each variable if there were no corner solutions, whereas the AMEs provide the mean marginal effect of each variable for the subjects in the sample, taking into account that some subjects are at the corner solutions. Thus, the AMEs are more informative as they can be interpreted as the marginal change from the treatments in the expected proportion of pro-environmental effort at the population level. Wooldridge (2010) refers to these as average partial effects, or average treatment effects when referring to dummy variables.

Table 4.5: Tobit models testing treatment effects

	Dependent variable:			
	Proportion pro-environmental effort			
	(1)		(2)	
	Coefs	AMEs	Coefs	AMEs
Low damage	-0.1178**	-0.0461***	-0.1572**	-0.0609**
	(0.0465)	(0.0178)	(0.0640)	(0.0239)
Low effort	0.0081	0.0033	-0.0478	-0.0189
	(0.0361)	(0.0145)	(0.0638)	(0.0249)
Income effect	-0.0292	-0.0116	-0.0168	-0.0067
	(0.0367)	(0.0145)	(0.0348)	(0.0139)
Choice	0.1442	0.0589	0.1935	0.0795
	(0.1494)	(0.0620)	(0.2146)	(0.0896)
Chose low	-0.0609	-0.0243	0.2014	0.0806
	(0.2899)	(0.1150)	(0.3023)	(0.1204)
Choice*Chose low	-0.3409**	-0.1254**	-0.4881**	-0.1715***
	(0.1543)	(0.0510)	(0.2147)	(0.0625)
TG B	-0.3814*	-0.1382*	-0.3971*	-0.1432**
	(0.2249)	(0.0714)	(0.2270)	(0.0709)
TG C	-0.0277	-0.0110	-0.0587	-0.0232
	(0.1872)	(0.0741)	(0.1889)	(0.0739)
TG D	0.0528	0.0213	-0.1891	-0.0727
	(0.2964)	(0.1204)	(0.3103)	(0.1141)
TG E	-0.1355	-0.0531	0.3227	0.1320
	(0.3365)	(0.1287)	(0.5824)	(0.2395)
Low damage*TG B			-0.0186	-0.0074
			(0.1015)	(0.0402)
Low damage*TG C			0.0668	0.0271
			(0.1111)	(0.0457)
Low effort*TG C			0.0276	0.0111
			(0.0973)	(0.0394)
Low effort*TG D			0.0943	0.0385
			(0.0865)	(0.0358)
Inc. effect*TG E			0.0294	0.0118
			(0.0596)	(0.0242)
Choice*TG E			-0.1092	-0.0425
			(0.2234)	(0.0840)
Chose low*TG E			-0.8444	-0.2729*
			(0.6576)	(0.1502)
Ch.*Ch. low*TG E			0.3444	0.1453
			(0.2318)	(0.0989)
Constant	0.2008		0.2335*	
	(0.1267)		(0.1243)	
$\hat{\sigma}$	0.6719***		0.6650***	
	(0.1060)		(0.1054)	
N		215		215
P-value		0.0010		0.0007
Pseudo r-squared		0.0148		0.0190
Pseudo log-lik.		-624.82		-622.18

Notes: TG abbreviates "Treatment group". Standard errors are clustered at the subject level and in parentheses. The delta-method is used to calculate standard errors for average marginal effects (AMEs). *p<0.1; **p<0.05; ***p<0.01.

the treatment groups, random or otherwise, as is standard with differences-in-differences models.[14]

Model (2) in Table 4.5 includes all relevant interactions between treatments and treatment groups, given which treatment groups received which treatment. Thus, these interactions allow me to test whether the treatment effects are also affected by any differences between the treatment groups. There are no significant coefficients on the treatment-treatment group interactions. Including these interactions does increase the absolute size of the low damage coefficient, and changes the low effort coefficient from positive to negative, but it remains not statistically significant. Another important test from this model is for any order effects as treatment groups A and B saw the high damage round first and low damage second, whereas treatment group C saw the rounds in the opposite order. The instructions and practice round were consistent with the first round that subjects were given to fully control for any order effects. There is no statistical significance on the coefficient of low damage*TG C, thus there is no evidence that order effects are a significant driver of the results. Given the lack of significance on any of the treatment and treatment group interaction coefficients, and using the AIC and BIC criteria, model (1) is chosen as the preferred model over model (2) for the analysis of the hypotheses.

At this point it is useful to briefly discuss the choice of the Tobit model. The Tobit model is used in this analysis as it is a corner solution model (Wooldridge, 2010) and a large number of the dependent observations are corner solutions on the unit interval (47% at 0 and 13% at 1 for proportion of pro-environmental effort). The Tobit model accounts for the corner solutions that subjects have arrived at by modelling a probability that a subject has chosen 0 or 1. Average marginal effects in the Tobit thus account for the non-linearities in the dependent variable, which are not accounted for by OLS (Wooldridge,

[14]The different systems used for recruitment for some sessions, as noted in Section 4.3.2, may lead to some minor differences between the treatment groups; otherwise there is no other systematic difference between the sessions. Specifically ORSEE was used for treatment groups A, C and D, whereas SONA was used for treatment groups B and E. SONA allows all eligible subjects in the subject pool to sign up to any experimental session, whereas ORSEE only allows a random subset of subjects who receive an email invitation to sign up to a session. Treatment group B does have a statistically significantly lower effort level than treatment group A, thus it is important to conduct the analysis with the treatment group controls. The recruitment differences are not expected to affect the results regarding the hypotheses; this assumption is confirmed by the results of model (2).

Table 4.6: Estimated elasticities.

Parameters	High to low damages	High to low dam., moral licensing	High to low effort	Income effect
$\Delta M/M$ (est.)	-0.1593^{**}	-0.4910^{**}	-0.2615^{***}	-0.0388
$\Delta\epsilon/\epsilon$	0.5	0.5		
$\Delta\phi/\phi$	-0.3333	-0.3333	0.3333	
$\Delta Y/Y$				0.3333
$-\eta_\epsilon(M)$ (est.)	**0.32** ($\pm$0.26)	**0.98** ($\pm$0.95)		
$\eta_\phi(M)$ (est.)	**0.48** ($\pm$0.39)	**1.47** ($\pm$1.42)	**−0.78** ($\pm$0.31)	
$\eta_Y(M)$ (est.)				**−0.12** ($\pm$0.29)

Notes: Elasticities are estimated from model (1) of Table 4.5, though high to low effort is adjusted to account for lower absolute effort being required in the low effort treatment. The parameters in the upper section of the table are chosen such that the elasticities of interest can be calculated; for example $\eta_\phi(M)$ is estimated by dividing $\Delta M/M$ by $\Delta\epsilon/\epsilon$, as per the standard definition of an elasticity. Standard errors are calculated using the delta-method. Brackets contain 95% confidence intervals. *p<0.1; **p<0.05; ***p<0.01.

2010). Nevertheless, OLS estimates are provided in the Appendix as a robustness check. There are no unexpected or particularly significant differences between the estimates from the Tobit and OLS models.

I also undertake a rough test for misspecification of the Tobit model. The test is to compare Tobit coefficients, divided by the estimated variance, with Probit coefficients for models estimated on dummies for 0 and greater than 0, and less than 1 and 1 (Wooldridge, 2010). The results of this exercise indicate no issues of misspecification. In terms of alternative models, the Cragg hurdle model (Wooldridge, 2010) was experimented with and provided no notable differences with the Tobit. Thus, the Tobit was favoured as the more parsimonious model.

Table 4.6 shows the estimated elasticities from the preferred model, model (1) of Table 4.5. The middle section of the table shows the values used to calculated the elasticities in the bottom section. The first row in the bottom section shows the negative of the energy efficiency elasticity of pro-environmental effort, $-\eta_\epsilon(M)$, which I defined in Section 4.3.1 as the behavioural rebound effect. This is positive and estimated as 0.32 for high to low damages and 0.98 for high to low damages with moral licensing. The numbers in brackets next to the elasticity estimate give the 95% confidence intervals.

The next elasticity in Table 4.6 is $\eta_\phi(M)$, the efficiency elasticity of pro-environmental effort, which relates only to how much damage reduction is achieved by an additional unit of pro-environmental effort. Given the efficiency of pro-environmental effort, ϕ, is a function of energy efficiency, ϵ, for the change from high to low damages, $\eta_\phi(M)$ is closely related to $-\eta_\epsilon(M)$ for the first two columns. However, the high to low effort treatment change is constructed so that ϕ increases while ϵ stays the same, hence only $\eta_\phi(M)$ can be estimated for this difference in effort between treatments. For high to low effort, $\eta_\phi(M)$ is estimated to be -0.78. Given there are only two optional additional letters in the low effort treatment instead of three, the measure of the proportion of pro-environmental effort is adjusted to account for the fact that completing one extra letter for every word is a 0.43 proportion of pro-environmental effort for the high damage treatment, but a 0.57 proportion of pro-environmental effort for the low effort treatment. This elasticity is thus calculated after adjusting proportion of pro-environmental effort to be an equivalent of absolute level of effort between treatments, by multiplying proportion of pro-environmental effort for the low effort treatment by 3/4. Finally, the income elasticity of pro-environmental effort $\eta_Y(M)$, is given for the high income treatment relative to the high damage treatment. It is small and not statistically significant, as per the estimate in model (1), Table 4.5.

Table 4.7 shows the proportion of pro-environmental effort, regressed on the responses to the survey, using a Tobit model. Only the high damage treatment is included for treatment groups that saw the high damage treatment for the first round. Only these observations are included as some of the survey questions were given in regards to the first round, thus it is important only to compare groups receiving the same treatment in the first round. Results on other treatments and rounds are consistent with these results.

Model (1) in Table 4.7 regresses just the norm belief variable on pro-environmental effort, and shows this variable is a strong driver of pro-environmental effort. The coefficient is highly significant and positive, with the belief that 1 extra letter is completed by others on average being associated with a 0.21 increase in proportion of pro-environmental effort. Model (2) shows the range of controls available being regressed on pro-environmental effort, without the norm belief variable. In this model, gender and reported levels of pro-environmental behaviours are significant explanatory variables. Model (3) includes

Table 4.7: Tobit models of proportion pro-environmental effort in high damage treatment, treatment groups A, B, D, E.

	Dependent variable: Proportion pro-environmental effort					
	(1)		(2)		(3)	
	Coefs	AMEs	Coefs	AMEs	Coefs	AMEs
Norm belief	0.4526***	0.2098***			0.4213***	0.1967***
	(0.0527)	(0.0152)			(0.0525)	(0.0174)
Letters			-0.0031	-0.0015	0.0003	0.0001
			(0.0027)	(0.0013)	(0.0022)	(0.0010)
Age			0.0196	0.0092	0.0121	0.0057
			(0.0173)	(0.0081)	(0.0135)	(0.0063)
Female			0.3507***	0.1644***	0.1404	0.0661
			(0.1134)	(0.0508)	(0.0898)	(0.0423)
Low income			-0.1699	-0.0828	-0.1537	-0.0745
			(0.1523)	(0.0764)	(0.1189)	(0.0596)
Australian			0.0387	0.0183	0.0088	0.0041
			(0.1413)	(0.0672)	(0.1112)	(0.0521)
Env. behav.			0.2733**	0.1285**	0.1378	0.0643
			(0.1234)	(0.0566)	(0.0962)	(0.0447)
NEP scale			-0.0344	-0.0162	0.0164	0.0076
			(0.1275)	(0.0599)	(0.1003)	(0.0468)
Env. org.			0.1841	0.0904	0.1110	0.0537
			(0.1650)	(0.0841)	(0.1289)	(0.0643)
Political party			0.4913*	0.2529	0.2505	0.1263
			(0.2938)	(0.1542)	(0.2258)	(0.1206)
Vote Labor			-0.3292	-0.1415*	-0.2900*	-0.1249*
			(0.2096)	(0.0795)	(0.1660)	(0.0642)
Vote Greens			0.2305	0.1142	0.2473	0.1233
			(0.2409)	(0.1239)	(0.1879)	(0.0980)
Vote other			-0.0722	-0.0331	-0.2580	-0.1081
			(0.3539)	(0.1583)	(0.2818)	(0.1038)
Vote unsure			0.0843	0.0394	-0.0474	-0.0221
			(0.1600)	(0.0742)	(0.1268)	(0.0594)
Constant	-0.3331***		-0.8751		-1.0714	
	(0.0828)		(0.8307)		(0.6590)	
$\hat{\sigma}$	0.2663***		0.3967***		0.2303***	
	(0.0469)		(0.0717)		(0.0406)	
N		167		167		167
Observations		167		167		167
P-value		0.0000		0.0099		0.0000
Pseudo r-squared		0.2441		0.0873		0.2989
Log-likelihood		-120.00		-144.91		-111.31

Notes: The delta-method is used to calculate standard errors for average marginal effects (AMEs). The "vote" dummy variables are relative to Vote Liberal. *p<0.1; **p<0.05; ***p<0.01.

all relevant variables. There is little change to the coefficient on the norm belief variable compared with model (1), whereas the controls change in size and statistical significance. Treatment group controls were tested, but were not statistically significant and did not change the overall results. Their inclusion also worsened model fit using the AIC and BIC criteria, and thus have not been included in Table 4.7.

4.4.3 Within subject hypotheses

The behavioural rebound effect

I now discuss the results as they pertain to hypothesis H1, that there is a positive behavioural rebound effect. If the difference in pro-environmental effort of the low damage treatment minus the high damage treatment is negative, then the behavioural rebound effect will be positive. Table 4.4 finds that this difference between treatments is negative at the 1% level. This result is also supported by the Tobit models estimated in Table 4.5, with the negative and statistically significant coefficient on the low damage treatment dummy, which is relative to the high damage treatment. Thus, the first main result, corresponding to hypothesis H1, is:

Result 1 There is a positive behavioural rebound effect.

Thus, I fail to reject hypothesis H1. The behavioural rebound effect for high to low damages is estimated to be 0.32, as shown in Table 4.6.

Efficiency elasticity of pro-environmental effort

Hypothesis H2 is that the efficiency elasticity of pro-environmental effort, when there is no associated change in energy efficiency, is positive. This hypothesis is tested through estimating the difference between the high damage treatment and the low effort treatment. There is no positive effect found in the non-parametric testing in Table 4.4, nor on the low effort coefficient in Table 4.5. The results presented in Table 4.6 for the estimated elasticity controls for the difference in absolute pro-environmental effort, rather than just proportion of pro-environmental, given low effort requires just two optional extra letters to be completed per word, rather than three. The following result is found:

Result 2 The efficiency elasticity of pro-environmental effort associated with an increase in the efficiency of pro-environmental effort with no change in energy efficiency is estimated to be -0.78.

Thus, I reject hypothesis H2. In fact, the increase in efficiency of pro-environmental effort, through an increase in ϕ, implemented by lowering the effort required to lower damage, is met with a strong reduction in pro-environmental effort. Taking into account the confidence interval, I cannot rule out an elasticity of -1. This finding is particularly interesting given, as also shown in Table 4.6, the efficiency elasticity of pro-environmental effort is positive when the change in efficiency of pro-environmental effort, ϕ, is due to a change in energy efficiency, ϵ.

Income effect

In terms of hypothesis H3, the non-parametric Mann-Whitney U test for income effect in Table 4.4 shows no evidence for an income effect forming a part of the behavioural rebound effect. There is also no statistically significant income effect estimated in Table 4.5. These results are a rejection of hypothesis H3.

The change in income between the high damage and high income treatments is one third; perhaps this is not a large enough change on the income side to have an effect on pro-environmental effort. Another test available is whether subject income level has any effect on their pro-environmental effort, though given subjects are students there is little variation in their personal income. The results in Table 4.7 again present no evidence for personal income having a positive effect on pro-environmental effort. Also included in Table 4.7 is whether the number of letters a subject completes in a round (which also impacts income earned for the experiment) has an impact on pro-environmental effort. The coefficient on that variable is not significantly different from 0. Thus, the following result is found:

Result 3 There is no evidence of a positive income effect on pro-environmental effort.

4.4.4 Between subject hypotheses

Pro-environmental effort and observables

I first discuss hypothesis H4, which relates social norms, demographics and environmental values to level of pro-environmental effort. Table 4.7 shows the main results for this hypothesis, with model (1) showing a strong link between pro-environmental effort and beliefs about social norms. Model (2) of Table 4.7 shows that females are more inclined to put in pro-environmental effort, as are people who report more pro-environmental behaviours in their daily life. Interestingly, pro-environmental orientation, as measured by the NEP scale, does not statistically predict pro-environmental effort when controlling for other variables. Australian citizen is another variable that could be hypothesised to positively predict pro-environmental effort, given the tree planting charity plants indigenous trees within the state of Victoria, but it is not statistically significant. However, model (3) includes all relevant variables and supports the overall finding:

Result 4 The strongest driver of pro-environmental effort is beliefs about social norms.

Including both social norms and other variables in model (3) leads to the size and significance of the female and environmental behaviour coefficients dropping away relative to model (2), which does not include beliefs about social norms. Norm beliefs are positively correlated both with being female and environmental behaviours and model (3) suggests that norm belief itself is the most powerful driver of pro-environmental behaviours.

One potential criticism of the norm belief variable is that subjects just chose the number closest to their level of effort, given there is no incentive to consider the question more deeply. To address this criticism, treatment groups B and E were incentivised for the norm belief question, being told they would earn $1 if they chose the average number of optional extra letters completed closest to the actual level in their session. There is no statistical difference between subjects incentivised for this question and subjects not incentivised ($p = 0.73$), thus the incentivised and non-incentivised treatment groups are pooled for Table 4.7.

Moral licensing

I now discuss the final set of coefficients not yet addressed in Table 4.5. These coefficients are for the choice treatment dummy, chose low and the interaction between choice and chose low.[15] Formally, to be consistent with hypothesis H5, the choice*chose low coefficient must be less than the low damage coefficient. Hence, I conduct a one-sided t test that the coefficient on choice*chose low < low damage. For both models, the difference is significant at the 10% level ($p = 0.087$), hence:

Result 5 There is evidence consistent with moral licensing.

Therefore, I fail to reject hypothesis H5. The 10% statistical significance level of the moral licensing behaviour is also consistent with the meta-study on moral licensing, Blanken et al. (2015). They find that moral licensing has a small to medium effect size, and thus for a 5% significance level with statistical power of 80% the study would need 165 subjects each in a control and moral licensing treatment group. This sample size is within the norm for a laboratory experiment, given the cost of the method, but evidence of moral licensing at the 10% level with the sample size in this study is consistent with Blanken et al.'s (2015) estimated effect size of moral licensing.

The results in Table 4.6 show that the behavioural rebound effect is estimated to be roughly three times the size under moral licensing, compared with exogenous technological change. At 98%, this is a large rebound, pushing costless endogenous technological change roughly at the level of backfire. However, this estimated behavioural rebound effect does have a large confidence interval associated with it, so neither a much lower moral licensing

[15] The baseline pro-environmental effort of subjects who chose high damages in the choice treatment are accounted for in the treatment group dummies for groups D and E.

effect nor backfire can be ruled out.

Finally, I test hypothesis H5a, that the moral licensing effect is larger for those with higher pro-environmental orientation. To test this hypothesis I re-estimate model (1) from Table 4.5 after separating the sample into two groups - those with a higher than the median pro-environmental orientation, according to the NEP measure, and those with less than or equal to the median pro-environmental orientation. The results are shown in Table 4.8.

Model (1) in Table 4.8 shows the results for those subjects with a equal to or stronger than median pro-environmental orientation. I conduct a one-sided t test that the choice*chose low interaction coefficient is less than the low damage coefficient. This is significant at the 5% level. Model (2) thus shows the results for those subjects with a less than median pro-environmental orientation. Conducting the same one-sided t test, the results are not statistically significant (with a p-value of 0.38). Hence:

Result 5a The moral licensing effect is larger for those with a higher pro-environmental orientation and is statistically insignificant for those with a lower pro-environmental orientation.

Therefore, I fail to reject hypothesis H5a. The results in Table 4.8 show a further interesting pattern: the other treatment effects are also different between the groups with low and high pro-environmental orientation. The low damage treatment has no statistical significance for those with high pro-environmental orientation, while it is larger for those with low pro-environmental orientation. This result suggests that the positive behavioural rebound effect is driven by those with lower pro-environmental orientation, as their pro-environmental effort is more sensitive to changing incentives. Additionally, there is a statistically significant negative income effect for those with high pro-environmental orientation. This latter result is something of a puzzle, though the effect size is not large.

To test whether Result 5a is driven by propensity to undertake pro-environmental effort, which seems to be driven mostly by beliefs about social norms, I conduct the

Table 4.8: Tobits testing treatment effects, separated by NEP measure.

	Dependent variable:			
	Proportion pro-environmental effort			
	NEP > median		NEP ≤ median	
	(1)		(2)	
	Coefs	AMEs	Coefs	AMEs
Low damage	-0.0588	-0.0279	-0.1883***	-0.0589***
	(0.0616)	(0.0288)	(0.0693)	(0.0216)
Low effort	0.0609	0.0296	-0.0462	-0.0148
	(0.0398)	(0.0195)	(0.0684)	(0.0216)
Income effect	-0.0818**	-0.0385**	0.0580	0.0189
	(0.0363)	(0.0161)	(0.0720)	(0.0235)
Choice	0.3877	0.1907	-0.0187	-0.0060
	(0.2665)	(0.1286)	(0.0682)	(0.0219)
Chose low	0.5063	0.2314*	-0.7383*	-0.2177**
	(0.3121)	(0.1255)	(0.4434)	(0.1109)
Choice*Chose low	-0.5440**	-0.2208***	-0.2345*	-0.0720*
	(0.2674)	(0.0844)	(0.1259)	(0.0368)
TG B	-0.6712**	-0.2597***	-0.1088	-0.0343
	(0.2951)	(0.0803)	(0.3461)	(0.1065)
TG C	-0.1537	-0.0715	-0.0162	-0.0052
	(0.2305)	(0.1030)	(0.3373)	(0.1082)
TG D	-0.7826**	-0.3016***	0.9967**	0.3458**
	(0.3480)	(0.0919)	(0.4724)	(0.1453)
TG E	-0.6620*	-0.2593**	0.2905	0.0957
	(0.3929)	(0.1113)	(0.4896)	(0.1615)
Constant	0.4244**		0.0177	
	(0.1800)		(0.1921)	
$\hat{\sigma}$	0.4348***		0.9619***	
	(0.0869)		(0.2305)	
N		102		113
Observations		306		339
P-value		0.0099		0.0079
Pseudo r-squared		0.0467		0.0375
Pseudo loglik		-279.62		-320.02

Notes: TG abbreviates "Treatment group". Standard errors are clustered at the subject level and in parentheses. The delta-method is used to calculate standard errors for average marginal effects (AMEs). *p<0.1; **p<0.05; ***p<0.01.

same exercise from Table 4.8 in Table C.1 in the Appendix using the pro-environmental behaviours variable.[16] The pattern is not the same; neither the group with low nor the group with high levels of reported environmental behaviours has a statistically significantly significant moral licensing effect.

4.5 Discussion

The results presented above provide evidence for a positive behavioural rebound effect, and a negative efficiency elasticity of pro-environmental effort when the change in efficiency of pro-environmental effort occurs without a change in energy efficiency. The lack of a statistically significant income effect on pro-environmental effort suggests that these findings are mostly driven by the substitution effect between private earnings and reducing environmental damages. The estimated size of the behavioural rebound effect is 32%, which is a similar size to the average direct rebound effect measured in the field (Gillingham et al., 2016; Sorrell et al., 2009). Thus, the behavioural rebound effect is shown to be significant, which suggests a need for further work into augmenting models of the rebound effect to include social norms and pro-environmental preferences. Indeed, the power of social norms in influencing pro-environmental behaviour is highlighted in this study, as it has been in previous research (Allcott, 2011).

The strength that a laboratory experiment brings to this research is the ability to cleanly identify effects that might be difficult or impossible to identify in the field, namely pure pro-environmental behaviours. In the field, behaviours that are *prima facie* pro-environmental may in reality be undertaken for a range of other benefits that they might bring the individual as well as the environmental benefits they provide. The task of identifying pure pro-environmental behaviours is not helped by the fact that technologies are imperfect substitutes. A car and a bicycle both provide a transport service, but with vastly different associated attributes such as comfort, speed, fitness benefits and environmental damages. While co-benefits of pro-environmental behaviours are important, such

[16] It is difficult to conduct the equivalent exercise using the norm belief variable, as most respondents chose 1 and as a discrete variable it does not allow easy separation of the subjects to roughly two equal sized groups of high and low norm beliefs.

as fitness and enjoyment, this novel experiment provides strong evidence that individuals respond not only to private incentives that change with changing technology, but also incentives to put in effort for the environment.

The laboratory setting also has the desirable feature of allowing a clean distinction to be drawn between exogenous and endogenous technological change in order to test for moral licensing. Thus, the evidence in favour of moral licensing presented here is compelling, given it is demonstrated with a costless and essentially irrelevant choice and on a relatively small sample size with which to investigate moral licensing.

Perhaps most important to the literature on moral licensing is Result 5a. It shows an even larger and more significant moral licensing effect on the more pro-environmentally orientated half of the sample, where the 102 subjects in the high pro-environmental orientation subsample represent less than a third of the sample size recommended by Blanken et al. (2015) to measure moral licensing. Furthermore, this finding does seem to be related solely to underlying pro-environmental orientation and not to revealed pro-environmental effort, as shown when comparing Table 4.8, which uses the NEP scale to separate the models, with Table C.1, which uses reported level of pro-environmental behaviour. It is a particularly interesting finding given that it is reported pro-environmental behaviour that has some predictive power on the underlying level of pro-environmental effort in a given round, unlike the NEP scale, as shown in Table 4.7.

Nevertheless, the unique environment created in the laboratory also requires some caveats on the estimates of the elasticities. One main limitation in this study is the salience of environmental damage being much higher than the real world, given clear environmental damages with real time feedback to subjects. High salience is likely to encourage a higher level of pro-environmental behaviour than would be observed outside of the laboratory (Schubert, 2017).

Another potential limitation is that subjects are aware that they are being observed. While the data collected are anonymised, this aspect of the laboratory environment may still influence subjects to act more morally than they would in a private setting (Levitt and List, 2007). This limitation may mean that the overall level of pro-environmental effort is higher in the laboratory than in the field, but the size of the behavioural rebound effect may be smaller if the result of this attribute is also a smaller behavioural response to changes in the level of environmental damages. The same reasoning applies to moral licensing; it may also be smaller in the laboratory than in the field. The countervailing force to this is the fact that the behavioural rebound effect and moral licensing may be stronger given the higher salience of pro-environmental behaviours in the laboratory. With these limitations in mind, the laboratory environment still provides evidence that the behavioural rebound effect is significant and important and can help guide further research in the field. Additionally, all subjects face the same conditions and therefore the experiment is internally consistent for the purpose of testing the hypotheses.

Future research building on the novel experimental design used in this paper can add in more aspects of the rebound effect to help build a better picture of the relative importance of the behavioural rebound effect in the overall rebound effect in energy use. Another aspect that could be tested in future is the importance of real time feedback on environmental damage along with the effect of having uncertain environmental damages. It would also be worth investigating the power of social norms further. Areas to investigate include how information or priming about social norms at the beginning of the experiment might affect pro-environmental effort, rather than just asking subjects about their beliefs about social norms in the post-experiment survey.

More generally, this paper provides an impetus for more research to determine the importance in the field of the behavioural rebound effect and moral licensing. Careful thought must be put into developing theory and research that allows the identification of the behavioural rebound effect in the field. A welfare analysis of the rebound effect including the behavioural rebound effect would be useful to help analyse policy interventions. Aspects of the behavioural rebound effect that could be investigated in the field include looking for more evidence of moral licensing and how it operates over the short and long term after the purchase of a durable good. Finally, it would be worthwhile testing policy interventions to encourage pro-environmental behaviours within the context of ongoing technological change, where these policies are likely to be welfare enhancing.

4.6 Conclusion

In this paper I present a novel laboratory experiment, which provides both the ability to measure the level of pro-environmental effort, given a private cost to that effort, and how that effort changes with changing incentives. I find a behavioural rebound effect of around 32% associated with an increase in energy efficiency, which suggests that changes in pro-environmental effort contribute to the overall rebound effect in energy use. The results also show that technological changes that make pro-environmental effort easier or more efficient, without changing energy efficiency, are unlikely to increase pro-environmental effort when accounting just for the environmental benefits. Additionally, I demonstrate the importance of beliefs about social norms for explaining level of pro-environmental effort. Finally, moral licensing increases the size of the behavioural rebound effect when technology change is endogenous, particularly for those individuals with a stronger pro-environmental orientation of their attitudes and beliefs.

It is worth considering a couple of examples to close, given the diversity of potential applications for the results of this paper. First, I return to the transport example for a final time. Those purchasing efficient cars for environmental reasons are subject to the direct rebound effect, the behavioural rebound effect and moral licensing. While these effects are unlikely to lead to backfire in the field, given most empirical evidence to date (Gillingham et al., 2016), the latter two effects could perhaps be reduced by information provision to continue to encourage pro-environmental behaviours, with a focus on social norms. On the other hand, the results show that improving the pro-environmental efficiency of a behaviour is unlikely to lead to an increase in that behaviour. One example of this for transport is adding an electric assist to bicycles. Do so is unlikely to encourage more cycling on environmental grounds, though it could still increase the level of cycling if it enhances other benefits, such as enjoyment.

Another example is reducing greenhouse gas emissions from power generation. The purchase of rooftop solar panels may be subject to moral licensing, whereas building solar farms to reduce the carbon emissions of grid electricity is essentially exogenous from a consumer point of view, and therefore is unlikely to be subject to moral licensing.

However, consumers choosing renewable energy options for their power provider are still potentially subject to moral licensing. Thus, policies to increase renewable energy in the power grid across all providers, along with continued energy conservation programs, may be more effective in reducing carbon emissions than subsidising rooftop solar or relying on consumer demand for renewable power.

This paper provides strong evidence that it is important to consider the effect of technology change on pro-environmental behaviours. On the one hand, many organisations – governmental and non-governmental – spend considerable resources encouraging pro-environmental behaviours. On the other hand, technology change that is encouraged by similar or the same organisations has the potential to discourage these behaviours when that technology change reduces the environmental impact of consumption and therefore reduces the efficiency of pro-environmental behaviours. If technology change makes sacrifices for the environment redundant then that negates the need to encourage pro-environmental behaviours. However, environmental policy challenges such achieving the large emission cuts required to meet global climate change targets over the next few decades requires many actions – both technology change and sacrifice of consumption. Thus, it is important to consider how technology change interacts with incentives for pro-environmental behaviours to ensure resources expended on behaviour change are allocated in an optimal way.

Chapter 5

Conclusion

The three papers within this thesis investigate how preferences and motivations influence behaviour, and how incentives can change behaviours in expected or unexpected ways. An experimental economics methodology is applied throughout, with the experiments being implemented in the applied contexts of environment and health. While each chapter is written to be able to stand alone as a complete paper, I make some final remarks and suggestions for future research here. I hope that these areas for future research not only spark further thought and interest in the reader, but also demonstrate the potential for future contributions within the field of experimental economics.

5.1 Final summaries and future research

5.1.1 Paper 1: Preferences for intrinsically risky attributes

We utilise data from a fully incentivised risk preferences experiment to test whether particular types of intrinsic risk are driving choices within a discrete choice experiment (DCE). Using this method we can identify that, for new sources of municipal water supply, supply risk is important for the choices of participants, and new technology risk is not important. Thus, we demonstrate the potential to leverage experimentally elicited preferences to deepen understanding of the choices people make in other domains, such as within a stated preferences DCE.

While we apply risk preferences to choices over sources of new water supply in a DCE, the method could be applied to other DCEs, as well as to other choices in the field. Logically, the applications are for situations where some of the underlying reasons behind choices, such as how much individuals care about certain types of risk, are not clear and not easily understood without introducing some additional information about those individuals. It is worth testing the method with other instances of intrinsic risk, such as for health risks. One appropriate intrinsic health risk is food poisoning risk, for example chicken, with the potential for salmonella and campylobacter, versus beef, with a lower risk of food poisoning. Finally, it would also be useful to investigate the limitations of the method, by adding to the literature around how much financial risk preferences can apply to decisions in other domains, and the advantages or disadvantages of other methods for measuring risk preferences.

5.1.2 Paper 2: Intrinsic motivation and health

We measure baseline intrinsic motivation using a real effort task in the laboratory, and find that it strongly predicts waist-to-height ratio in the field. We also utilise our baseline measure of intrinsic motivation to help better understand the effects of a range of temporary incentives, both before and after their application. We find, on average, that the "pay – but do not pay too much" rule (of Pokorny, 2008) prevails. However, when we separate the groups into high and low motivation, we find that "pay – but do not pay too much" holds more closely for high motivation individuals, and "pay enough or don't pay at all" (of Gneezy and Rustichini, 2000b) is a better rule for low motivation individuals. On average, the high power monetary incentive is most likely to crowd out intrinsic motivation after its removal, which is important when considering temporary incentives.

The result that intrinsic motivation predicts waist-to-height ratio is striking, and warrants further research. An initial question to investigate is: what other behaviours and outcomes, health or otherwise, can underlying intrinsic motivation predict in the field? A second and related area for research is into what refinements can be made to the measure of intrinsic motivation. Alternative tasks could be tested, to test to what extent the measure of intrinsic motivation depends on the task. Finally, that intrinsic motivation is found to be a predictor of health outcomes demonstrates the importance of understanding intrinsic motivation when it comes to promoting healthy behaviours, and compliance with treatment programmes. Building on insights from this and other similar studies, it would be valuable to continue research into how best to increase healthy behaviours in the field.

5.1.3 Paper 3: A behavioural rebound effect

In this paper, I design a laboratory experiment to test for a behavioural rebound effect and moral licensing. The experimental design allows for the isolation of a rebound effect in pure pro-environmental behaviours. It isolates these behaviours from other behaviours that are ostensibly pro-environmental, but that are undertaken for other reasons, such as to save money. I find that there is indeed a behavioural rebound effect, which increases in size when moral licensing also occurs. Moral licensing is most pronounced for more environmentally orientated people.

This paper lends itself to future research in both the laboratory and the field. First, the novel experimental design can be used for future research around pro-environmental preferences, environmental policies and the rebound effect. With regards to environmental policies, the experimental design could be used for testing various price and non-price policy interventions to encourage pro-environmental behaviours. This future research could even take the form of testing a range of interventions and their interactions, from taxes and subsidies, to nudges using social norms or an increased saliency of environmental damages. In terms of the rebound effect, it is worth exploring how the experiment could be augmented to investigate a full rebound effect. In this vein, it is worth testing whether there is any interaction between a decrease in environmental benefit and an increase in private benefits when it comes to the size of the rebound effect.

In the field, the evidence for a behavioural rebound effect and the evidence for moral licensing provides motivation for further research. It is worth exploring methods to tease out the relative size of the behavioural rebound effect versus the direct rebound effect. It is also worth exploring evidence for moral licensing further, and if there is evidence for it, testing for whether it operates primarily over the short term, or whether it persists over the longer term. The applications for the paper could be investigated further too. If possible, it would be interesting to test whether subsidised solar for household is indeed worse for the behavioural rebound effect than exogenous increases in renewable energy provided through the electricity grid.

Finally, the welfare implications need to be teased out further. To what extent is the behavioural rebound effect of concern, and is it likely that future progress in environmentally friendly technology might offset the losses in pro-environmental effort? Given the enormous environmental challenges faced around the world on the one hand, and the rapid progress being made towards improving the environmental performance of technology on the other, these questions are worth answering.

Bibliography

Abbott, A., S. Nandeibam, and L. O'Shea (2013). Recycling: Social norms and warm-glow revisited. *Ecological Economics 90*, 10–18.

Abeler, J. and D. Nosenzo (2015). Self-selection into laboratory experiments: Pro-social motives versus monetary incentives. *Experimental Economics 18*(2), 195–214.

Adamowicz, W., J. Louviere, and M. Williams (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management 26*, 271–292.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics 95*(9-10), 1082–1095.

Allcott, H. and S. Mullainathan (2010). Behavior and energy policy. *Science 327*(5970), 1204–1205.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383–405.

Anderson, L. R. and J. M. Mellor (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics 27*(5), 1260–1274.

Andersson, H., A. R. Hole, and M. Svensson (2016). Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety. *Journal of Environmental Economics and Management 75*, 41–53.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior and Organization 116*, 451–464.

Andreoni, J. and C. Sprenger (2012). Risk preferences are not time preferences. *American Economic Review 102*(7), 3357–3376.

Araujo, F. A., E. Carbone, L. Conell-Price, M. W. Dunietz, A. Jaroszewicz, R. Landsman, D. Lam, L. Vesterlund, S. W. Wang, and A. J. Wilson (2016). The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association 2*(1), 1–12.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–555.

Ashwell, M. and S. D. Hsieh (2005). Six reasons why the waist-to-height ratio is a rapid and effective global indicator for health risks of obesity and how its use could simplify the international public health message on obesity. *International Journal of Food Sciences and Nutrition 56*(5), 303–307.

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics 130*(3), 1067–1115.

Australian Bureau of Statistics (2011). Victoria (main statistical area structure) census data.

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

Bartczak, A., S. Chilton, and J. Meyerhoff (2015). Wildfires in Poland: The impact of risk preferences and loss aversion on environmental choices. *Ecological Economics 116*, 300–309.

Bénabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics 126*(2), 805–855.

Benndorf, V., H. A. Rau, and C. Sölch (2014). Minimizing learning behavior in experiments with repeated real-effort tasks. *Available at SSRN 2503029*.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*, 122–142.

Bertrand, M. and S. Mullainathan (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review 91*(2), 67–72.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. *Review of Economics and Statistics 94*(3), 686–698.

Blanken, I., N. van de Ven, and M. Zeelenberg (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin 41*(4), 540–558.

Bolle, F. and P. E. Otto (2010). A price is a signal: On intrinsic motivation, crowding-out, and crowding-in. *Kyklos 63*(1), 9–22.

Botzen, W. and J. van den Bergh (2012). Risk attitudes to low-probability climate change risks: WTP for flood insurance. *Journal of Economic Behavior and Organization 82*(1), 151–166.

Botzen, W. J. W. and J. C. Van Den Bergh (2012). Monetary valuation of insurance against flood risk under climate change. *International Economic Review 53*(3), 1005–1026.

Bowles, S. and S. Polanía-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, 368–425.

Bradford, D., C. Courtemanche, G. Heutel, P. McAlvanah, and C. Ruhm (2014). Time preferences and consumer behavior. Technical report, National Bureau of Economic Research.

Bradford, W. D. (2010). The association between individual time preferences and health maintenance habits. *Medical Decision Making 30*(1), 99–112.

Brekke, K. A., S. Kverndokk, and K. Nyborg (2003). An economic model of moral motivation. *Journal of Public Economics 87*(9-10), 1967–1983.

Brent, D. A., L. Friesen, L. Gangadharan, and A. Leibbrandt (2017). Behavioral insights from field experiments in environmental economics. *International Review of Environmental and Resource Economics 10*(2), 95–143.

Brent, D. A., L. Gangadharan, A. Leroux, and P. A. Raschky (2014). Putting one's money where one's mouth is: Increasing saliency in the field. Discussion Paper 43/14, Monash University, Melbourne, Australia.

Brüggen, A. and M. Strobel (2007). Real effort versus chosen effort in experiments. *Economics Letters 96*(2), 232–236.

Cameron, J., K. M. Banko, and W. D. Pierce (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *Behavior Analyst 24*(1), 1.

Cameron, T. A. and J. DeShazo (2013). Demand for health risk reductions. *Journal of Environmental Economics and Management 65*(1), 87–109.

Carson, R. T. and J. J. Louviere (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics 49*(4), 539–559.

Cason, T. N. and L. Gangadharan (2002). Environmental labeling and incomplete consumer information in laboratory markets. *Journal of Environmental Economics and Management 43*(1), 113 – 134.

Cavalcanti, C., S. Engel, and A. Leibbrandt (2013). Social integration, participation, and community resource management. *Journal of Environmental Economics and Management 65*(2), 262–276.

Cerasoli, C. P., J. M. Nicklin, and M. T. Ford (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin 140*(4), 980–1008.

Chan, N. W. and K. Gillingham (2015). The microeconomic theory of the rebound effect and its welfare implications. *Journal of the Association of Environmental and Resource Economists 2*(1), 133–159.

Chang, J.-j., W.-N. Wang, and J.-Y. Shieh (2017). Environmental rebounds/backfires: Macroeconomic implications for the promotion of environmentally-friendly products. *Journal of Environmental Economics and Management*.

Chapman, G. B. and E. J. Coups (1999). Time preferences and preventive health behavior: Acceptance of the influenza vaccine. *Medical Decision Making 19*(3), 307–314.

Charness, G. and U. Gneezy (2009). Incentives to exercise. *Econometrica 77*(3), 909–931.

Charness, G., U. Gneezy, and A. Imas (2013). Experimental methods: eliciting risk preferences. *Journal of Economic Behavior and Organization 87*, 43–51.

Chen, Z., H. H. Ngo, W. Guo, X. C. Wang, C. Miechel, N. Corby, A. Listowski, and K. O'Halloran (2013). Analysis of social attitude to the new end use of recycled water for household laundry in Australia by the regression models. *Journal of Environmental Management 126*, 79–84.

Costa, D. L. and M. E. Kahn (2013). Energy conservation nudges and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*(3), 680–702.

Croson, R. and N. Treich (2014). Behavioral environmental economics: Promises and challenges. *Environmental and Resource Economics 58*(3), 335–351.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty 41*(3), 219–243.

De Borger, B., I. Mulalic, and J. Rouwendal (2016). Measuring the rebound effect with micro data: A first difference approach. *Journal of Environmental Economics and Management 79*, 1–17.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology 18*(1), 105–115.

Deci, E. L., R. Koestner, and R. M. Ryan (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin 125*(6), 627–668.

Deci, E. L., R. M. Ryan, and R. Koestner (2001). The pervasive negative effects of rewards on intrinsic motivation: Response to Cameron (2001). *Review of Educational Research 71*(1), 43–51.

DEFRA (2008). A framework for pro-environmental behaviours. Technical report, Department for Environment, Food and Rural Affairs, London.

DellaVigna, S. and D. Pope (2017). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*.

Dinda, S. (2004). Environmental Kuznets curve hypothesis: A survey. *Ecological Economics 49*(4), 431–455.

Dolnicar, S., A. Hurlimann, and B. Grün (2014). Branding water. *Water Research 57*, 325–338.

Duarte, R., V. Pinilla, and A. Serrano (2014). Looking backward to look forward: Water use and economic growth from a long-term perspective. *Applied Economics 46*(2), 212–224.

Duke, J. M., S. J. Dundas, R. J. Johnston, and K. D. Messer (2014). Prioritizing payment for environmental services: Using nonmarket benefits and costs for optimal selection. *Ecological Economics 105*, 319–329.

Dunlap, R. E., K. D. Van Liere, A. G. Mertig, and R. E. Jones (2000). Measuring endorsement of the New Ecological Paradigm: A revised NEP scale. *Journal of Social Issues 56*(3), 425–442.

Erkal, N., L. Gangadharan, and B. H. Koh (2017). Monetary and non-monetary incentives in real-effort tournaments. *European Economic Review*.

Erkal, N., L. Gangadharan, and N. Nikiforakis (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review 101*(7), 3330–3348.

Fehr, E. and A. Leibbrandt (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics 95*(9-10), 1144–1155.

Ferraro, P. J. and M. K. Price (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics 95*(1), 64–73.

Fielding, K. S., J. Gardner, Z. Leviston, and J. Price (2015). Comparing public perceptions of alternative water sources for potable use: The case of rainwater, stormwater, desalinated water, and recycled water. *Water Resources Management 29*(12), 4501–4518.

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

Frey, B. S. and F. Oberholzer-Gee (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 746–755.

Friesen, L. and L. Gangadharan (2013). Environmental markets: What do we learn from the lab? *Journal of Economic Surveys 27*(3), 515–535.

Ghosh, N. K. and M. F. Blackhurst (2014). Energy savings and the rebound effect with multiple energy services and efficiency correlation. *Ecological Economics 105*, 55–66.

Gill, D. and V. L. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Gillingham, K., D. Rapson, and G. Wagner (2016). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy 10*(1), 68–88.

Glenk, K. and S. Colombo (2013). Modelling outcome-related risk in choice experiments. *Australian Journal of Agricultural and Resource Economics 57*(4), 559–578.

Global Commission on the Economy and Climate (2014). Better growth better climate: The new climate economy report. Technical report, New Climate Economy.

Gneezy, U., A. Leibbrandt, and J. A. List (2016). Ode to the sea: Workplace organizations and norms of cooperation. *Economic Journal 126*(595), 1856–1883.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives 25*(4), 191–210.

Gneezy, U. and A. Rustichini (2000a). A fine is a price. *Journal of Legal Studies 29*(1), 1–17.

Gneezy, U. and A. Rustichini (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*(3), 791–810.

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014). Adolescent time preferences predict lifetime outcomes. *Economic Journal 124*, F739–F761.

Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology 93*(1), 48–58.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Halvorsen, B. (2008). Effects of norms and opportunity cost of time on household recycling. *Land Economics 84*(3), 501–516.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström (2005). Risk aversion and incentive effects: Comment. *American Economic Review 95*(3), 897–901.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics 109*(2), 341–368.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009–1055.

Hawcroft, L. J. and T. L. Milfont (2010). The use (and abuse) of the New Environmental Paradigm scale over the last 30 years: A meta-analysis. *Journal of Environmental Psychology 30*(2), 143–158.

Hensher, D., N. Shore, and K. Train (2005). Households willingness to pay for water service attributes. *Environmental and Resource Economics 32*(4), 509–531.

Hensher, D. A. and W. H. Greene (2003). The mixed logit model: The state of practice. *Transportation 30*(2), 133–176.

Heyman, J. and D. Ariely (2004). Effort for payment: A tale of two markets. *Psychological Science 15*(11), 787–793.

Hidrue, M. K., G. R. Parsons, W. Kempton, and M. P. Gardner (2011). Willingness to pay for electric vehicles and their attributes. *Resource and Energy Economics 33*(3), 686–705.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Hunt, L. C. and D. L. Ryan (2014). Catching on the rebound: Why price elasticities are generally inappropriate measures of rebound effects. Technical report, Surrey Energy Economics Centre (SEEC), School of Economics, University of Surrey. Available at https://ideas.repec.org/p/sur/seedps/148.html.

Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics 114*, 14–18.

Jevons, W. (1865). *The coal question: An inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*. Macmillan.

Johnston, R. J., S. K. Swallow, T. J. Tyrrell, and D. M. Bauer (2003). Rural amenity values and length of residency. *American Journal of Agricultural Economics 85*(4), 1000–1015.

Kajackaite, A. and P. Werner (2015). The incentive effects of performance requirements – A real effort experiment. *Journal of Economic Psychology 49*, 84–94.

Kamenica, E. (2012). Behavioral economics and psychology of incentives. *Annual Review of Economics 4*(1), 427–452.

Kerr, J., M. Vardhan, and R. Jindal (2012). Prosocial behavior and incentives: Evidence from field experiments in rural Mexico and Tanzania. *Ecological Economics 73*, 220–227.

Khalil, E. L. (1997). Buridan's ass, risk, uncertainty, and self-competition: A theory of entrepreneurship. *Kyklos 50*(2), 147–163.

Lacetera, N., M. Macis, and R. Slonim (2014). Rewarding volunteers: A field experiment. *Management Science 60*(5), 1107–1129.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy 74*(2), 132–157.

Lee, J. and U. Hwang (2016). Hypothetical bias in risk preferences as a driver of hypothetical bias in willingness to pay: Experimental evidence. *Environmental and Resource Economics 65*(4), 789–811.

Levitt, S. D. and J. A. List (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives 21*(2), 153–174.

List, J. A. and J. F. Shogren (1998). Calibration of the difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior and Organization 37*(2), 193–205.

Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in China. *Review of Economics and Statistics 95*(4), 1386–1403.

Louviere, J. J., T. N. Flynn, and R. T. Carson (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling 3*(3), 57–72.

Lusk, J. L. and K. H. Coble (2005). Risk perceptions, risk preference, and acceptance of risky food. *American Journal of Agricultural Economics 87*(2), 393–405.

Ma, Q., J. Jin, L. Meng, and Q. Shen (2014). The dark side of monetary incentive: How does extrinsic reward crowd out intrinsic motivation. *NeuroReport 25*(3), 194–198.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Economics*, pp. 105–142. New York: Academic Press.

Meas, T., W. Hu, M. T. Batte, T. A. Woods, and S. Ernst (2015). Substitutes or complements? Consumer preference for local and organic food attributes. *American Journal of Agricultural Economics 97*(4), 1044–1071.

Melton, B. E., W. E. Huffman, J. F. Shogren, and J. A. Fox (1996). Consumer preferences for fresh food items with multiple quality attributes: Evidence from an experimental auction of pork chops. *American Journal of Agricultural Economics 78*(4), 916–923.

Miao, B. and S. Zhong (2015). Comment on "Risk preferences are not time preferences": Separating risk and time preference. *American Economic Review 105*(7), 2272–2286.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, and A. Caspi (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences 108*(7), 2693–2698.

Monin, B. and D. T. Miller (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology 81*(1), 33.

Newell, R. G. and J. Siikamäki (2014). Nudging energy efficiency behavior: The role of information labels. *Journal of the Association of Environmental and Resource Economists 1*(4), 555–598.

Newell, R. G. and J. Siikamäki (2015). Individual time preferences and energy efficiency. *American Economic Review 105*(5), 196–200.

Nyborg, K., R. B. Howarth, and K. A. Brekke (2006). Green consumers and public policy: On socially contingent moral motivation. *Resource and Energy Economics 28*(4), 351–366.

Olmstead, S. M. (2010). The economics of managing scarce water resources. *Review of Environmental Economics and Policy 4*(2), 179–198.

Petrolia, D. R., C. E. Landry, and K. H. Coble (2013). Risk preferences, risk perceptions and flood insurance. *Land Economics 89*(2), 227–245.

Pindyck, R. S. (2007). Uncertainty in environmental economics. *Review of Environmental Economics and Policy 1*(1), 45–65.

Pokorny, K. (2008). Pay – but do not pay too much. *Journal of Economic Behavior and Organization 66*(2), 251–264.

Productivity Commission (2011). Australia's urban water sector: Productivity Commission inquiry report volume 1. Inquiry Report 55, Canberra.

Promberger, M. and T. M. Marteau (2013). When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. *Health Psychology 32*(9), 950–957.

Qiu, Y., G. Colson, and C. Grebitus (2014). Risk preferences and purchase of energy-efficient technologies in the residential sector. *Ecological Economics 107*, 216–229.

Roberts, D. C., T. A. Boyer, and J. L. Lusk (2008). Preferences for environmental quality under uncertainty. *Ecological Economics 66*(4), 584–593.

Rolfe, J. and J. Windle (2015). Do respondents adjust their expected utility in the presence of an outcome certainty attribute in a choice experiment? *Environmental and Resource Economics 60*(1), 125–142.

Royer, H., M. Stehr, and J. Sydnor (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics 7*(3), 51–84.

Scarpa, R., R. Zanoli, V. Bruschi, and S. Naspetti (2012). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics 95*(1), 165–180.

Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics 132*, 329–342.

Shao, J. and R. R. Sitter (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association 91*(435), 1278.

Smith, A. (1759). *The theory of moral sentiments*. Andrew Millar, in the Strand; and Alexander Kincaid and J. Bell, in Edinburgh.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London: W. Strahan and T. Cadell.

Sorrell, S. and J. Dimitropoulos (2008). The rebound effect: Microeconomic definitions, limitations and extensions. *Ecological Economics 65*, 636–649.

Sorrell, S., J. Dimitropoulos, and M. Sommerville (2009). Empirical estimates of the direct rebound effect: A review. *Energy Policy 37*(4), 1356–1371.

Sturm, B. and J. Weimann (2006). Experiments in environmental economics and some close relatives. *Journal of Economic Surveys 20*(3), 419–457.

Takahashi, H., J. Shen, and K. Ogawa (2016). An experimental examination of compensation schemes and level of effort in differentiated tasks. *Journal of Behavioral and Experimental Economics 61*, 12–19.

Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving decisions about health, wealth, and happiness* (Revised and expanded ed.). New York: Penguin Books.

Tiefenbeck, V., T. Staake, K. Roth, and O. Sachs (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy 57*, 160–171.

Tonin, M. and M. Vlassopoulos (2014). Corporate philanthropy and productivity: Evidence from an online real effort experiment. *Management Science 61*(8), 1795–1811.

Train, K. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge; New York: Cambridge University Press.

Viceisza, A. C. G. (2016). Creating a lab in the field: Economics experiments for policy-making. *Journal of Economic Surveys 30*(5), 835–854.

Whitehead, J. C., S. K. Pattanayak, G. L. Van Houtven, and B. R. Gelso (2008). Combining revealed and stated preference data to estimate the nonmarket value of ecological services: An assessment of the state of the science. *Journal of Economic Surveys 22*(5), 872–908.

Wielgus, J., L. R. Gerber, E. Sala, and J. Bennett (2009). Including risk in stated-preference economic valuations: Experiments on choices for marine recreation. *Journal of Environmental Management 90*(11), 3401–3409.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, Mass: MIT Press.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

Appendix A

Appendix for Chapter 2

A.1 Overall DCE choices

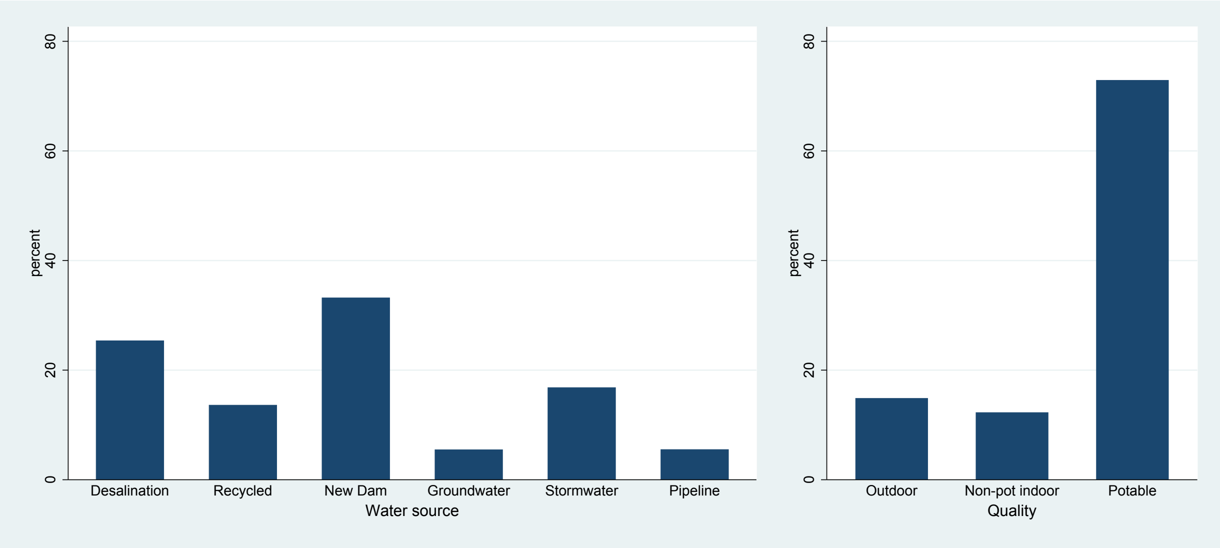
Figure A.1 shows the overall results from the DCE, with new dam and desalination being the most preferred options, and groundwater and pipeline the least preferred. It is important to remember that desalination and new dam always had potable water, whereas the other four water sources had a balanced mix of allowed use (quality) levels. Therefore, if ensuring water is potable is a concern for individuals, then desalination and new dam never had to be ruled out on the basis of allowed use. The rightmost section of Figure A.1 shows the aggregate choices for allowed use, regardless of cost and water source. Potable is by far the most popular allowed use at 72.9%, followed by non-potable outdoor (14.9%) and non-potable indoor (12.3%).

A.2 Imputing risk attitudes for the full sample

Table A.1 displays the tobit model that is used to impute risk attitudes for the full sample. The fitted values from this model are used to impute the risk attitudes for those without observations for this variable. In order to more accurately impute risk attitudes, both demographics and indicators of attitudes to risk are included.[1] The attitude to

[1] The model is estimated from 124 of the 137 people with observed risk attitudes as the other 13 do not have a full set of right-hand side variables due to answering "Don't know" or refusing to answer to some of the survey questions.

Figure A.1: Overall percentage of choices made by participants.



risk variables are flood risk perception, owning flood insurance, not knowing whether or not they own flood insurance, and an interaction between owning flood insurance and flood risk perception. The flood risk perception question is shown in Table 2.3. In the tobit model it is treated as a Likert-type scale from 1 to 5, with 1 equating to "Almost never" and 5 being "1 in 2 years". As already mentioned, the locations chosen for the survey had similar rainfall patterns, so differences in responses should not be a reflection of differences in actual flood risk; rather they should reflect differences in perceived flood risk. The interaction between owning flood insurance and flood risk perception is positive and statistically significant, as expected.

The first demographics included in Table A.1 are age, gender and education. Next are dummies for middle and high household income (relative to low income) as self-identified by participants. This variable is used for income as subjective data can be useful as explanatory variables to explain behaviour (Bertrand and Mullainathan, 2001). Furthermore, more people were willing to answer this question about their household income than giving a more precise indication in dollar values. Finally, the dummy variables for the council areas of Fairfield, Moonee Valley and Manningham are included, and are relative to Warringah. The differences in risk attitudes by location likely reflect the different mix of ethnicities and cultural backgrounds, owing to immigration patterns, of

Table A.1: Tobit for imputing coefficient of CRRA.

	Tobit
Constant	-0.4976
	(0.7710)
Flood risk perception	-0.0916
	(0.1109)
Own flood insurance	-0.0225
	(0.2435)
Don't know flood insurance	-0.3901
	(0.2498)
Flood insurance*Flood risk percep	0.3412**
	(0.1503)
Age	-0.0089
	(0.0061)
Female	0.0810
	(0.1840)
Education (yrs)	0.0591
	(0.0452)
Middle income	-0.2087
	(0.2376)
High income	-0.1150
	(0.3547)
Fairfield	0.6196**
	(0.2870)
Moonee Valley	0.3039
	(0.2558)
Manningham	0.6336**
	(0.2644)
σ	0.9600***
	(0.0729)
Pseudo R-squared	0.0798
P-value	0.0047
N	124

Notes: Standard errors are in parentheses. Middle and high income are dummies relative to low income. Dummies for Fairfield, Moonee Valley and Manningham are relative to Warringah. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the different council areas.

As shown in the last rows of Table A.1, the model overall has a good statistical fit. Even if most of the coefficients are not individually significant, the low p-value of 0.005 for the full model shows that they have a high level of joint significance.

We also include Table A.2 in this appendix to replicate Table 2.5, but estimated using just the 137 individuals for whom risk preferences are observed. The overall results between the two tables are similar, but with overall a lower level of statistical significance on the coefficients in Table A.2 as expected.

Table A.2: Mixed logit regression results - those with observed risk preference data only.

	Base	All with risk	Supply Risk	Technology Risk
	(1)	(2)	(3)	(4)
Fixed Coefficients & Means				
Fixed Coefficients				
Non-potable outdoor	-0.0587	-0.0587	-0.0583	-0.0587
	(0.1080)	(0.1081)	(0.1080)	(0.1080)
Non-potable indoor	-0.2602^{**}	-0.2599^{**}	-0.2595^{**}	-0.2601^{**}
	(0.1202)	(0.1203)	(0.1202)	(0.1201)
$\beta_{r,desalination}$		-0.9913^{**}		
		(0.4417)		
$\beta_{r,recycled}$		-0.1791		
		(0.4311)		
$\beta_{r,groundwater}$		0.3353		
		(0.4380)		
$\beta_{r,stormwater}$		-0.1140		
		(0.3410)		
$\beta_{r,pipeline}$		-0.1826		
		(0.3490)		
$\beta_{r,supply}$			0.2247	
			(0.2729)	
$\beta_{r,tech}$				-0.1209
				(0.2759)
Random Coefficients				
Desalination	-0.6763^{***}	0.3287	-0.4354	-0.6748^{***}
	(0.1781)	(0.4697)	(0.3400)	(0.1783)
Recycled	-1.3025^{***}	-1.1034^{**}	-1.0638^{***}	-1.1771^{***}
	(0.2077)	(0.5041)	(0.3551)	(0.3528)
Groundwater	-2.1674^{***}	-2.5231^{***}	-1.9428^{***}	-2.1664^{***}
	(0.2540)	(0.5479)	(0.3707)	(0.2538)
Stormwater	-0.7050^{***}	-0.5866	-0.7071^{***}	-0.5778^{*}
	(0.1558)	(0.3911)	(0.1558)	(0.3283)
Pipeline	-1.6346^{***}	-1.4344^{***}	-1.6346^{***}	-1.6342^{***}
	(0.1774)	(0.4089)	(0.1771)	(0.1774)
Cost	-0.1538	-0.1662	-0.1543	-0.1528
	(0.1108)	(0.1020)	(0.1107)	(0.1112)
Standard Deviation or Spread				
Standard Deviation				
Desalination	1.6336^{***}	1.5527^{***}	1.6085^{***}	1.6400^{***}
	(0.1987)	(0.1783)	(0.2009)	(0.1991)
Recycled	1.6001^{***}	1.6775^{***}	1.6101^{***}	1.5955^{***}
	(0.2123)	(0.2246)	(0.2144)	(0.2099)
Groundwater	1.2644^{***}	1.2648^{***}	1.2825^{***}	1.2624^{***}
	(0.2344)	(0.2343)	(0.2359)	(0.2339)
Stormwater	1.1651^{***}	1.1628^{***}	1.1648^{***}	1.1628^{***}
	(0.1433)	(0.1398)	(0.1428)	(0.1431)
Pipeline	0.7929^{***}	0.7856^{***}	0.7937^{***}	0.7921^{***}
	(0.1945)	(0.1942)	(0.1925)	(0.1942)
Spread				
Cost	0.2485	0.2828	0.2502	0.2462
	(0.2702)	(0.2349)	(0.2700)	(0.2719)
AIC	4128.2	4132.7	4129.5	4130.0
BIC	4201.3	4231.9	4207.8	4208.3
Observations	1370	1370	1370	1370
Individuals	137	137	137	137

Note: Standard errors clustered at the respondent level are in parentheses. The coefficient for cost follows a triangular distribution. All other random coefficients are normally distributed. Allowed use variables are relative to potable, water source variables are relative to new dam. All models are estimated using 500 Halton draws. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A.3 Instructions - incentivised risk task

-------------------------------- [NEW SCREEN] --------------------------------

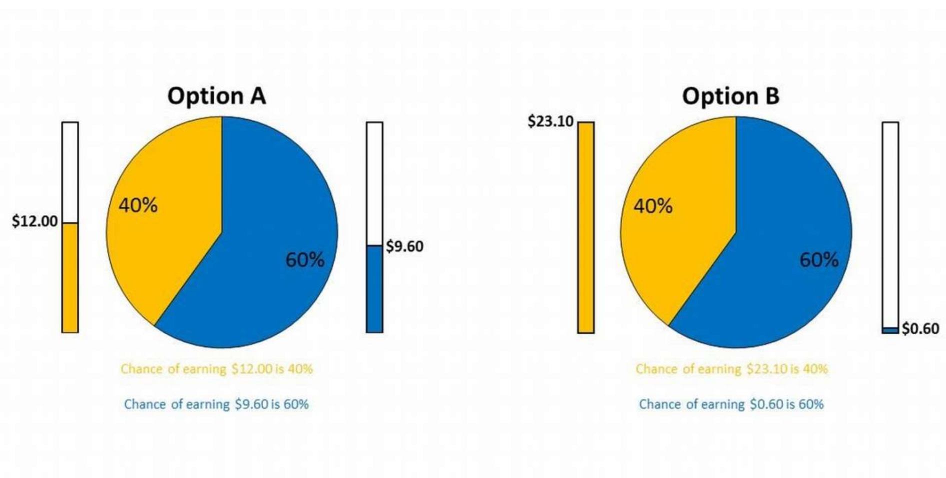
ACTIVITY 1

Explanation

Water management in Australia is influenced by weather and many other uncertain factors. Therefore, as a first step, we would like to get a better understanding how Australians make decisions related to uncertainty. There are standard techniques to make responses comparable between individual respondents. We are using one of these techniques here, to understand how important uncertainty is to you, by asking you to make a series of 10 choices in simple decision problems, in which you will earn some money. How much you receive will depend partly on **chance** and partly on the **choices** you make. The decision problems are not designed to test you. The only right answer is what you really would choose.

For each decision problem, please state whether you prefer option A or option B. After answering all 10 decision problem, **one of the 10** decision problems will be randomly selected and its chance outcome will be given to you as payment. As any of the decisions can be chosen for payment, you should pay attention to the choice you make in every decision screen.

Example1a: Here is an example of one choice that you may see on the screen.



- If Option A was chosen, there is a 40% chance that you will be paid $12.00 and a 60% chance that you will be paid $9.60.
- If Option B was chosen, there is a 40% chance that you will be paid $23.10 and a 60% chance that you will be paid $0.60.

Example1b: Here is an example of one choice that you may see on the screen.

- If Option A was chosen, there is a 70% chance that you will be paid $12.00 and a 30% chance that you will be paid $9.60.
- If Option B was chosen, there is a 70% chance that you will be paid $23.10 and a 30% chance that you will be paid $0.60.

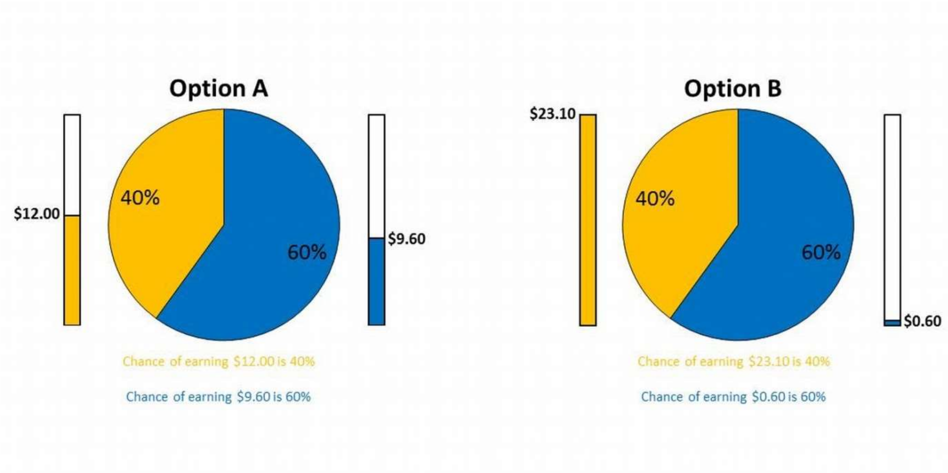
In short, this activity is trying to explore how you respond to risk.

How will you be paid?

As previously mentioned prior to the examples, you will earn some money depending on **choices** you made, and through **chance**.

After you have completed the 10 decision problems for this activity you will be shown a random number generator where you will be prompted to click “Stop!!” button. The generated random number will determine which of the 10 decision problems to focus on. If the random generator number was a 7, the “decision problem” to focus on will the 7th shown decision problem.

After a random number has been generated, you will be asked to draw another random number through the random number generator. The second random number generator will determine how much you will earn.



Referring back to the earlier examples, we mentioned the scenario below.

- If Option A was chosen, there is a 40% chance that you will be paid $12.00 and a 60% chance that you will be paid $9.60.
- If Option B was chosen, there is a 40% chance that you will be paid $23.10 and a 60% chance that you will be paid $0.60.

If in the above example, you had chosen Option A, and the number drawn from the second random number generator was between 1 and 4, then you earn $12.00. If the number drawn was between 5 and10, then you earn $9.60.

If in the above example, you had chosen Option B, and the number drawn from the second random number generator was between 1 and 4, then you earn $23.10. If the number drawn was between 5 and10, then you earn $0.60.

All earnings are in cash and are in addition to the $30 initial endowment that you receive as compensation for your time and effort in this and the following parts of this study. The interviewer will pay you the final balance of your earnings when all parts of the study are completed.

PLEASE TAKE IN TO CONSIDERATION THAT THERE ARE NO CORRECT OR WRONG DECISIONS. WE ARE ONLY TRYING TO EXPLORE DEPENDING ON THE DECISION PROBLEMS GIVEN HOW YOU RESPOND TO RISK.

A.4 Instructions - discrete choice experiment.

------------------------------- [NEW SCREEN] -------------------------------

ACTIVITY 2

When water shortages become more frequent, investments to increase urban water supply need to be made. There are a number of options in terms of water source and technology that a city can invest in. These options differ with respect to the quality of water provided and therefore their allowed use, as well as the cost of water provision. It is possible to install a third water pipe to your house, so that your tap water will not be contaminated with potentially lower quality water from the new source. You would **NOT** have to pay for the installation of the third pipe.

You will now be asked to make a series of 10 choices regarding your preferred additional water source, its allowed uses and the resulting cost of water. Assume that this would be the cost of your total water consumption per kilolitre in AUD. No other rates or charges would change.

PLEASE TAKE IN TO CONSIDERATION THAT THERE ARE NO CORRECT OR WRONG DECISIONS. THESE DECISION PROBLEMS ARE NOT DESIGNED TO TEST YOU AND YOUR RESPONSE WILL NOT RESULT IN YOU PAYING MORE FOR YOUR WATERBILL.

[USE INSTRUCTIONS CHOICE SET 2 HERE AND EXPLAIN DIFFERENT ATTRIBUTE LEVELS]

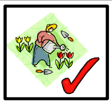
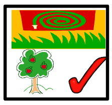
Example 2: Here is an example of one choice set that you may see on the screen.

You can choose between one of the six additional water sources. If the water from your preferred source is not supplied at drinking water quality, assume that a third water line has been installed to your home at no additional cost other than the new water price per kl of water you consume.

Do you have any questions?

Figure A.2: Information sheet provided for participants of discrete choice experiment.

Source of Additional Water	**Explanation**
Desalinisation	Desalinated sea water
Recycled	Recycled grey water
New Dam	Water drawn from new dam in the catchment
Groundwater	Water drawn from underground aquifers
Stormwater	Locally harvested and treated stormwater by your council
Pipeline	Water transported via pipelines from outside the catchment, for example from rural areas.

Allowed uses	**Explanation**	**Levels**
Note that: 1) any additional water sourced may be used to water non-edible garden plants (Class D).	**Limited Outdoor** water receives lowest treatment of all classes. May only be used for non-food garden plants (**ornamental plants and flowers, no lawns**). Other outdoor uses that do not involve human contact permitted. (Class D)	
2) Technological solutions exist to bring water from any source up the highest (fit for drinking) standard.	**Outdoor**: in addition to the **limited outdoor** uses listed above, outdoor water may be used to **water lawns and fruit trees** grown over a meter high (Classes B and C).	
	Indoor water may be used for **all outdoor** uses (including the watering of vegetable gardens) and for **limited indoor**, including for clothes washing and closed system toilet flushing.	
	Potable water is of drinking quality and **allowed for any use**.	
Price/kl	This is the price you would be charged per kilolitre of your total, billed water consumption.	$1.60-$3.20

Appendix B

Appendix for Chapter 3

B.1 Balance Test between treatment groups

Table B.1 shows a multinomial logit balance test between treatments on the main covariates of the subjects, where the control group is the base category. Overall, the model is not statistically significant; only two coefficients of the 36 are significant, which does not suggest any systematic differences between the treatment groups.

Table B.1: Multinomial logit balance test between treatments.

	Treatment group (relative to control):			
	Low power	High power	High power thresh	Charity
Constant	−0.765	−1.623	2.329	0.543
	(2.930)	(3.017)	(2.844)	(2.958)
Age	−0.007	−0.019	−0.002	−0.006
	(0.016)	(0.017)	(0.016)	(0.017)
Female	−0.464	−0.022	−0.557	0.081
	(0.438)	(0.447)	(0.426)	(0.445)
Education	0.029	0.053	−0.031	−0.016
	(0.153)	(0.158)	(0.149)	(0.155)
Personal income	−0.004	0.006	−0.002	0.004
	(0.008)	(0.007)	(0.008)	(0.008)
Impatience	−0.049	−0.006	−0.072	0.094
	(0.089)	(0.090)	(0.085)	(0.089)
Present bias	0.818	0.788	0.601	0.481
	(0.584)	(0.597)	(0.548)	(0.608)
Future bias	1.202**	1.119**	0.020	0.531
	(0.535)	(0.538)	(0.575)	(0.534)
Waist-to-height ratio	1.452	1.766	−2.174	−1.936
	(3.047)	(3.044)	(3.149)	(3.087)
Observations				229
Log Likelihood				−355.097
LR test p-value				0.757

Notes: Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

B.2 Experimental instructions

Figure B.1: Overview, practice round and round 1 instructions, part 1 (same for all treatment groups).

Instructions

Overview

Please note: all instructions will be read aloud by the experimenter.

Please switch off all cell phones until you leave this room.

Toilets are available at the end of the corridors by the lifts. Male toilets are on this side and female are on the other side. In the event of an emergency we will evacuate the building. Emergency exit stairs are located at the end of the corridor on the far side of the lifts.

As part of today's experiment, you will be participating in four activities: Activity 1, Activity 2, Activity 3 and Activity 4. Each activity may consist of one or more rounds. You will receive detailed instructions about each of the activities before you participate in them. We anticipate the total time for the experiment to be roughly one and a half hours.

You have been assigned a random ID number so that everything you do today cannot be identified as coming from you. Your ID number is overturned on your desk – please keep this here until the end of the experiment.

As indicated in the explanatory statement, you may earn money for participating in some of the activities. Earning details will be explained at the start of the activity. An administrative assistant will be paying you at the end of the laboratory session in the neighbouring room. He or she will not be involved in analysing the data from this experiment. I will record the payment details for each ID number and hand this to an administrative assistant. At the end of the session you will be instructed to take the ID number on your desk, and hand this to an administrative assistant, who will organise your payment.

After you have completed all the activities, we would like you to answer some questions about yourself. Please take your time and answer honestly and as accurately as possible.

We are about to begin the first activity. Please listen carefully. We will explain the activity and then you will have a chance to practice. Do not talk or discuss the activity with people around you. There will be opportunities to ask questions to be sure that you understand how to perform each activity. At any time during this experiment, please wait at your seat and do not do anything unless instructed by the experimenter. Also, do not look at other's responses at any time during this experiment.

Figure B.2: Overview, practice round and round 1 instructions, part 2 (same for all treatment groups).

ACTIVITY 1 Instructions

Activity 1 consists of a word encoding task. The task involves correctly assigning numbers to 5 random letters that you are given. First, the task will be explained in detail. Then you will be given 2 minutes to practise.

Explanation:

Near the top of each screen you will be given the full alphabet. Below each letter will be a number. Here is an example of part of what you will see:

A	B	C	D	E	F	G	H	I	J	K	L	M	N
11	25	14	12	6	17	24	2	22	15	10	18	21	7

…

In the centre of the screen you will see five randomly selected letters. This is your "word". Below each letter is an empty box, as shown here:

WORD:	P	C	G	S	N
CODE:					

In order to encode the "word" you can click on each box with your mouse and type the number associated with each letter. You may only use the numbers at the top of your keyboard, and *not* the numbers in the number pad at the side of your keyboard.

After you have completed a word it will be counted if you click OK with your mouse. The OK button is located at the bottom of the screen. If you click OK you will be given a new word to encode. The computer will not give you a new word until the word you have encoded is correct.

We are about to begin a 2 minute practice round. The screen will appear exactly as it will in the actual activity. Note that the number of seconds remaining for the round is displayed at top right of the screen. The number of words you have encoded is displayed below the alphabet.

If you have any questions about this task please raise your hand now.

The practice round will begin shortly.

After the practice round:

You will now be given the task for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.3: Round 2 instructions, control.

ACTIVITY 1, Round 2 Instructions

You will now be given the same task again.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.4: Round 2 instructions, low power incentive (same as high power, other than the piece rate values).

ACTIVITY 1, Round 2 Instructions

You will now be given the same task again. However, this time you will be paid 5c for every word you have correctly completed. You will receive any earnings you make at the end of the experiment from an administrative assistant.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.5: Round 2 instructions, high power with threshold incentive.

ACTIVITY 1, Round 2 Instructions

You will now be given the same task again. However, this time you will be paid $23 if you complete 23 words. Above 23 words, you will be paid $1 for every additional word. You will receive any earnings you make at the end of the experiment from an administrative assistant.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.6: Round 2 instructions, charity incentive.

ACTIVITY 1, Round 2 Instructions

You will now be given the same task again. However, this time every 2 words you complete will fund the planting of one indigenous tree in Victoria. A local environmental charity will receive the funds to plant these trees after the experiment. To prove this money has been donated, in the coming days you will be sent an email with a receipt stating the total amount donated to the charity from everyone in this session. This information will include the average number of trees that will be planted per person from this session.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.7: Round 3 instructions, control.

ACTIVITY 1, Round 3 Instructions

You will now be given the same task again.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.8: Round 3 instructions, monetary incentives.

ACTIVITY 1, Round 3 Instructions

You will now be given the same task again. This time you will not be paid.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.9: Round 3 instructions, charity incentive.

ACTIVITY 1, Round 3 Instructions

You will now be given the same task again. This time you will not be funding tree planting.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.10: Time preferences task instructions.

ACTIVITY 2 Instructions

This next task is designed to help us understand how you trade off monetary payments over time. You will be making decisions involving real money. Therefore, it is important that you carefully consider your answers.

You will see two sets of questions. In the first set, you will be making choices between receiving an amount of money today or a larger amount of money in 5 weeks' time. In the second set of questions you will also be asked whether you want to receive an amount of money in 5 weeks, or a larger amount in 10 weeks' time.

Here is an example from the first set of questions. You must click the circle for Option A or Option B:

	Option A			Option B
Decision 4:	$10 today	○ A	○ B	$11 in 5 weeks

From the two sets of questions, you will be asked a total 18 questions. One of your 18 answers will be randomly chosen by the computer, and you will be paid for this answer only.

Let's say the computer randomly chose to pay you for your answer for the question above.

Let's say you chose Option A. You will be given a $10 WISH voucher at the end of the experiment today. This will be in addition to any cash payments you have earned today.

Let's say you chose Option B. You will be mailed an $11 WISH voucher to reach you in 5 weeks' time. Only the administrative assistant will have access to your name and address details – you will remain anonymous to the researchers. She will take every step she can to have your voucher arrive at your address as close to 5 weeks from today as possible.

Remember, one of the answers to the 18 questions will affect your final payment today. Therefore, you should answer truthfully for every question which of the two options you prefer.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.11: Round 4 instructions, control.

ACTIVITY 3 Instructions

You will now be given the word encoding task from Activity 1 again.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.12: Round 4 instructions, monetary incentives.

ACTIVITY 3 Instructions

You will now be given the word encoding task from Activity 1 again. You will not be paid.

The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

Figure B.13: Round 4 instructions, charity incentive.

ACTIVITY 3 Instructions

You will now be given the word encoding task from Activity 1 again. You will not be funding tree planting.

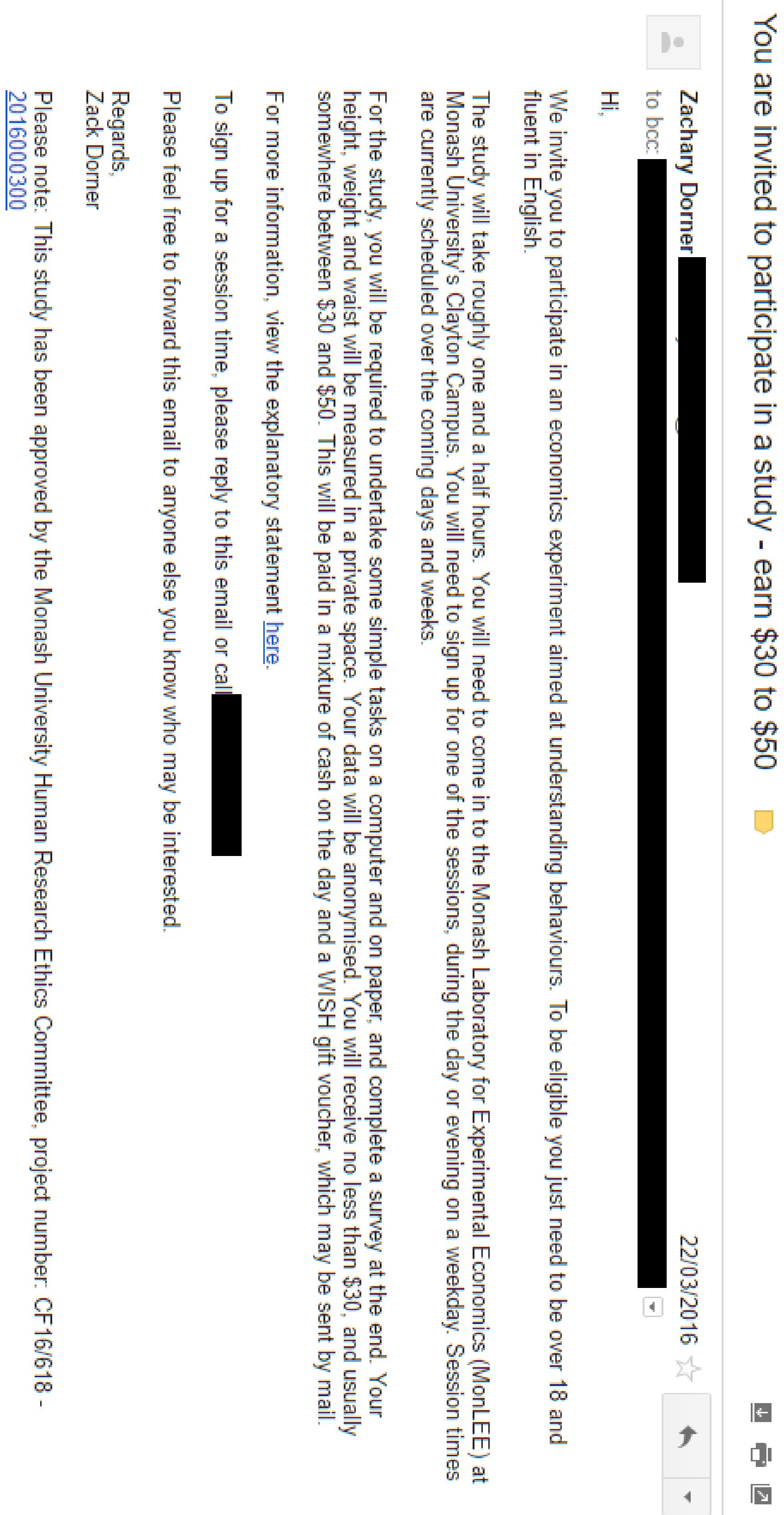
The task will run for 5 minutes.

If you have any questions please raise your hand now.

The task will begin shortly.

B.3 Recruitment email and advertisement examples

Figure B.14: Recruitment email.



You are invited to participate in a study - earn $30 to $50

Zachary Dorner [redacted] 22/03/2016

to bcc: [redacted]

Hi,

We invite you to participate in an economics experiment aimed at understanding behaviours. To be eligible you just need to be over 18 and fluent in English.

The study will take roughly one and a half hours. You will need to come in to the Monash Laboratory for Experimental Economics (MonLEE) at Monash University's Clayton Campus. You will need to sign up for one of the sessions, during the day or evening on a weekday. Session times are currently scheduled over the coming days and weeks.

For the study, you will be required to undertake some simple tasks on a computer and on paper, and complete a survey at the end. Your height, weight and waist will be measured in a private space. Your data will be anonymised. You will receive no less than $30, and usually somewhere between $30 and $50. This will be paid in a mixture of cash on the day and a WISH gift voucher, which may be sent by mail.

For more information, view the explanatory statement here.

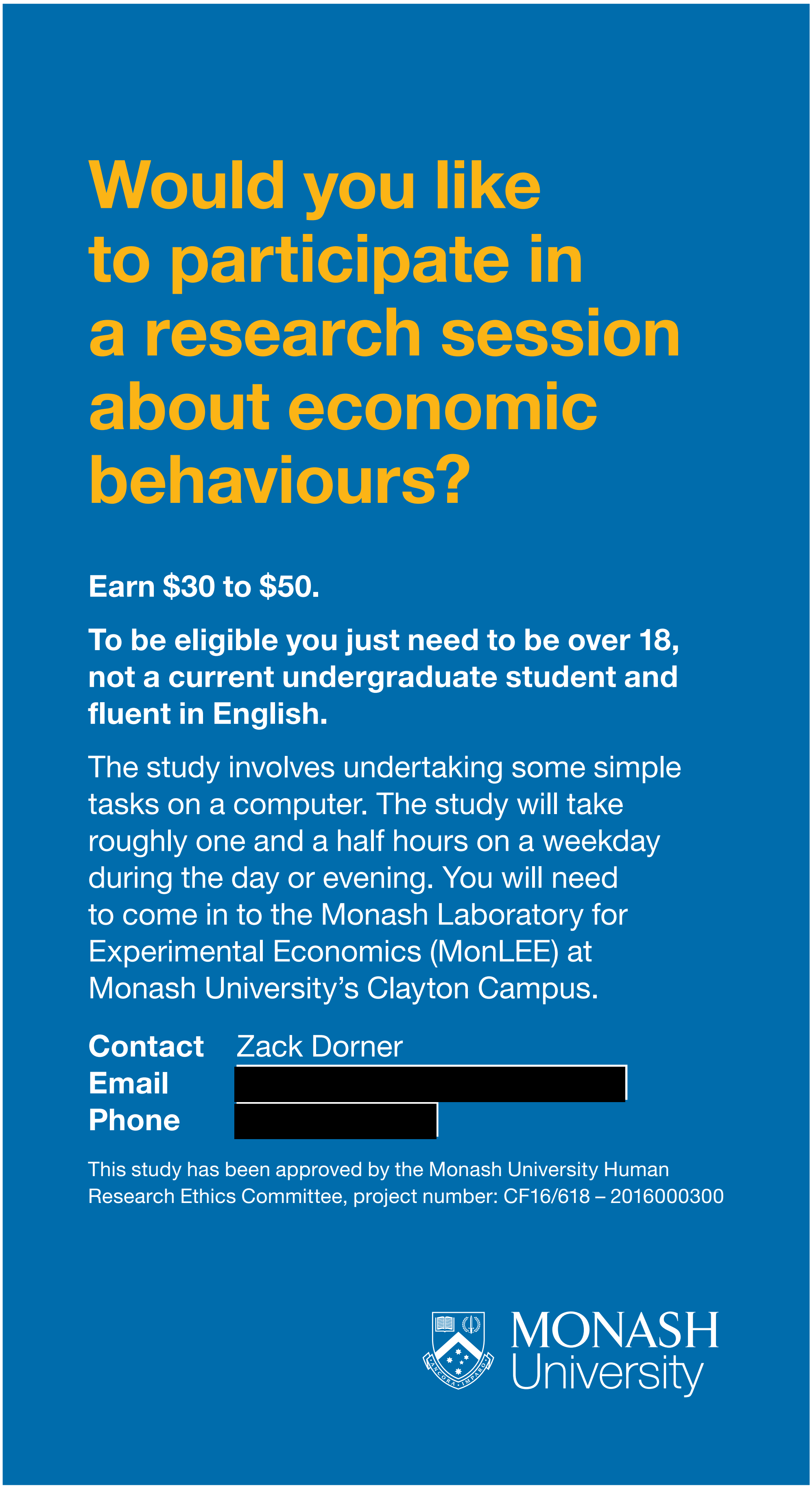
To sign up for a session time, please reply to this email or call [redacted].

Please feel free to forward this email to anyone else you know who may be interested.

Regards,
Zack Dorner

Please note: This study has been approved by the Monash University Human Research Ethics Committee, project number: CF16/618 - 2016000300

Figure B.15: Recruitment advertisement.



Appendix C

Appendix for Chapter 4

C.1 Supplementary results

Table C.1: Tobits testing treatment effects, separated by environmental behaviours.

	Dependent variable:			
	Proportion pro-environmental effort			
	Envi beh. > median		Envi beh. ≤ median	
	(1)		(2)	
	Coefs	AMEs	Coefs	AMEs
Low damage	-0.0655*	-0.0305*	-0.1857*	-0.0598**
	(0.0347)	(0.0162)	(0.0953)	(0.0296)
Low effort	-0.0290	-0.0136	0.0487	0.0164
	(0.0522)	(0.0242)	(0.0513)	(0.0173)
Income effect	-0.0269	-0.0126	-0.0283	-0.0094
	(0.0341)	(0.0159)	(0.0758)	(0.0250)
Choice	-0.0247	-0.0116	0.2803	0.0980
	(0.0560)	(0.0261)	(0.2692)	(0.0972)
Chose low	0.3454	0.1591	-0.4239	-0.1352
	(0.3488)	(0.1516)	(0.4333)	(0.1299)
Choice*Chose low	-0.1507**	-0.0690**	-0.5137*	-0.1496**
	(0.0743)	(0.0334)	(0.2797)	(0.0682)
TG B	-0.5374**	-0.2189**	-0.1815	-0.0580
	(0.2678)	(0.0859)	(0.3709)	(0.1130)
TG C	-0.0345	-0.0161	0.0113	0.0038
	(0.2379)	(0.1106)	(0.2937)	(0.0982)
TG D	-0.4996	-0.2111*	0.6355	0.2302
	(0.3464)	(0.1217)	(0.4704)	(0.1717)
TG E	-0.5171	-0.2159	0.1562	0.0530
	(0.4110)	(0.1413)	(0.4978)	(0.1708)
Constant	0.3870**		-0.0297	
	(0.1521)		(0.2147)	
$\hat{\sigma}$	0.4964***		0.8617***	
	(0.1022)		(0.2053)	
N		101		114
Observations		303		342
P-value		0.0087		0.0399
Pseudo r-squared		0.0268		0.0304
Pseudo loglik		-290.42		-318.91

Notes: TG abbreviates "Treatment group". Standard errors are clustered at the subject level and in parentheses. The delta-method is used to calculate standard errors for average marginal effects (AMEs). *p<0.1; **p<0.05; ***p<0.01.

Table C.2: OLS testing treatment effects as in Table 4.5.

	Dependent variable:	
	Prop. pro-environmental effort	
	(1)	(2)
Low damage	-0.0501***	-0.0499**
	(0.0169)	(0.0213)
Low effort	0.0083	-0.0069
	(0.0159)	(0.0283)
Income effect	-0.0090	-0.0278
	(0.0129)	(0.0194)
Choice	0.0564	0.0614
	(0.0622)	(0.0886)
Chose low	-0.0439	0.0691
	(0.1236)	(0.1262)
Choice*Chose low	-0.1311**	-0.1548*
	(0.0627)	(0.0851)
TG B	-0.1548*	-0.1379
	(0.0843)	(0.0926)
TG C	-0.0096	-0.0185
	(0.0810)	(0.0855)
TG D	0.0041	-0.0944
	(0.1261)	(0.1275)
TG E	-0.0459	0.1757
	(0.1437)	(0.2511)
Low damage*TG B		-0.0471
		(0.0404)
Low damage*TG C		0.0113
		(0.0385)
Low effort*TG C		0.0154
		(0.0420)
Low effort*TG D		0.0281
		(0.0378)
Inc. effect*TG E		0.0298
		(0.0262)
Choice*TG E		0.0045
		(0.1033)
Chose low*TG E		-0.3681
		(0.2798)
Ch.*Ch. low*TG E		0.0422
		(0.1020)
Constant	0.3704***	0.3754***
	(0.0575)	(0.0596)
N	215	215
P-value	0.0000	0.0000
Ajd. r-squared	0.0119	0.0081

Notes: TG abbreviates "Treatment group". Standard errors are clustered at the subject level and in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table C.3: OLS of proportion pro-environmental effort in High damage treatment, treatment groups A, B, D, E as in Table 4.7.

	Dependent variable:		
	Proportion pro-environmental effort		
	(1)	(2)	(3)
Norm belief	0.2481***		0.2299***
	(0.0245)		(0.0257)
Letters		-0.0011	0.0004
		(0.0014)	(0.0012)
Age		0.0065	0.0016
		(0.0091)	(0.0074)
Female		0.1638***	0.0690
		(0.0576)	(0.0480)
Low income		-0.1209	-0.1111*
		(0.0796)	(0.0647)
Australian		0.0555	0.0375
		(0.0742)	(0.0603)
Env. behav.		0.1419**	0.0775
		(0.0653)	(0.0536)
NEP scale		-0.0849	-0.0533
		(0.0670)	(0.0546)
Env. org.		0.1464*	0.1066
		(0.0883)	(0.0719)
Political party		0.3386**	0.2061
		(0.1623)	(0.1327)
Vote Labor		-0.1177	-0.1090
		(0.1052)	(0.0855)
Vote Greens		0.1272	0.1407
		(0.1282)	(0.1042)
Vote other		-0.0707	-0.1508
		(0.1912)	(0.1556)
Vote unsure		0.0788	0.0065
		(0.0854)	(0.0698)
Constant	0.0534	0.0409	-0.0693
	(0.0347)	(0.4293)	(0.3490)
N	167	167	167
Observations	167	167	167
P-value	0.0000	0.0049	0.0000
Adj. r-squared	0.3803	0.1017	0.4069

Notes: The "vote" dummy variables are relative to Vote Liberal. *p<0.1; **p<0.05; ***p<0.01.

Table C.4: OLS testing treatment effects, separated by NEP and environmental behaviours, as in Tables 4.8 and C.1.

	Dependent variable: Proportion pro-environmental effort			
	NEP > median	NEP $\leq$ median	E beh. > median	E beh. $\leq$ median
	(1)	(2)	(3)	(4)
Low damage	-0.0363	-0.0625***	-0.0311**	-0.0670**
	(0.0275)	(0.0208)	(0.0153)	(0.0288)
Low effort	0.0331	-0.0127	-0.0069	0.0214
	(0.0203)	(0.0245)	(0.0280)	(0.0169)
Income effect	-0.0185	-0.0036	-0.0118	-0.0054
	(0.0138)	(0.0211)	(0.0175)	(0.0190)
Choice	0.1579	-0.0217	-0.0116	0.0955
	(0.1141)	(0.0491)	(0.0427)	(0.0931)
Chose low	0.1432	-0.2516	0.1769	-0.1947
	(0.1021)	(0.1711)	(0.1268)	(0.1683)
Choice*Chose low	-0.2252**	-0.0604	-0.0792*	-0.1549*
	(0.1130)	(0.0543)	(0.0466)	(0.0933)
TG B	-0.2933**	-0.0538	-0.2498**	-0.0681
	(0.1301)	(0.1170)	(0.1157)	(0.1216)
TG C	-0.0615	-0.0096	-0.0043	-0.0074
	(0.1280)	(0.1136)	(0.1214)	(0.1084)
TG D	-0.3277**	0.3263*	-0.2465**	0.1974
	(0.1308)	(0.1786)	(0.1232)	(0.1806)
TG E	-0.2400	0.1216	-0.2552	0.0823
	(0.1682)	(0.1857)	(0.1692)	(0.1894)
Constant	0.4502***	0.3290***	0.4284***	0.3148***
	(0.1053)	(0.0675)	(0.0824)	(0.0802)
N	102	113	101	114
Observations	306	339	303	342
P-value	0.0001	0.001	0.0002	0.0339
Adj. r-squared	.0429	.0313	.0191	.0188

Notes: TG abbreviates "Treatment group". Standard errors are clustered at the subject level and in parentheses. *p<0.1; **p<0.05; ***p<0.01.

C.2 Experimental instructions and survey questions

What follows is the experimental instructions given to treatment groups B and E. Treatment groups A, C and D did not have an incentivised question for the survey. Treatment group C was given the low damage parameters in the example and practice round as they were given this treatment first. Otherwise the instructions are identical between treatment groups.

Figure C.1: Overview instructions.

Instructions

Overview

Please note: all instructions will be read aloud by the experimenter.

Please turn off all cell phones for the duration of this experiment.

As part of today's experiment, you will be participating in an activity and a survey. The activity will consist of 3 rounds. There will also be a practice round at the start of the activity. You will receive detailed instructions before the practice round and before the actual rounds. We anticipate the total time for the experiment to be roughly one and a half hours.

As part of today's experiment, you will be earning money. You will be paid based on your effort and the decisions you make from one of the 3 rounds in the activity. The computer will randomly choose the round for which you will be paid. In addition to any earnings you might have in these tasks, you will be given $10 for participating in today's experiment. Any money that you earn in the experiment will be paid to you in cash at the end of the experiment. All payments will be rounded to the nearest 5c.

After you have completed the activity, we would like you to answer some questions about yourself in the survey. Please take your time and answer honestly and as accurately as possible. You will not be identified in any way. Your survey answers will only be used for this experiment and will only be used by the researcher(s) involved in this project.

We are about to begin the activity. Please listen carefully. It is important that you understand the instructions of the task properly. If you do not understand, you will not be able to participate effectively. The task will be explained with examples. Do not talk or discuss the activity with people around you. There will be opportunities to ask questions at the end of each set of instructions if you are unsure how to perform the activity. At any time during this experiment, please remain in your seat and follow the instructions of the experimenter. Also, do not look at any other person's responses at any time during this experiment.

Figure C.2: Practice round instructions part 1.

PRACTICE ROUND Instructions

The activity consists of a word encoding task. The task involves correctly assigning numbers to up to 9 random letters that you are given. First, the task will be explained in detail. You will have the opportunity to ask questions. Then you will be given a practice round, followed by another opportunity to ask questions. Finally, you will be given the task for 3 rounds. Each round will last 8 minutes.

Explanation:

Near the bottom of each screen you will be given the full alphabet in a random order. Below each letter will be a number. Here is an example of what you will see:

D	P	Q	E	G	O	U	I	T	S	R	V	F	N	B	C	H	W	A	M	Y	X	Z	K	L	J
59	28	25	91	87	11	20	26	49	34	61	35	99	85	82	63	96	83	38	76	88	67	29	46	17	72

In the centre of the screen, to the left, you will see 6 randomly selected letters. This is your "word". Below each letter is an empty box, as shown here:

WORD :	E	B	J	K	O	M
CODE:						

In order to encode the "word" it is necessary to click on each box with your mouse and type the number associated with each letter. You may only use the numbers at the top of your keyboard, and *not* the numbers in the number pad at the side of your keyboard.

After you have completed a word it will be counted if you click OK with your mouse. The OK button is located at the bottom of the screen. If you click OK you will be given a new word to encode. The computer will not give you a new word until the word you have encoded is correct.

You will earn money for each word you successfully encode. The amount you have earned so far in a round will be displayed at the top of your screen.

For the practice round, it has been set so that you "earn" 60c for each word you encode. However, as this is a practice round, you will **not** actually be paid this money. Your earnings will be paid to you at the end of the experiment for one out of the 3 actual rounds. The round that will be paid out will be decided at random by the computer, so you need to treat all non-practice rounds as if they may be the one that will affect your final earnings. The one round that is paid out will be the same for everyone in this session.

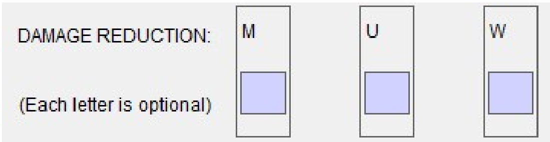
In addition to the earnings you make in a round, there will also be an amount donation to a charity. This amount starts at \$336 for the full session per round, which is \$14 per person per round. The one round that is chosen to be paid out for will also be the only round for which the donation to the charity is made.

Every word you encode will lower the charity payment. The damages per word to the charity has been set at 54c per word for the practice round. Again, this is for demonstration purposes only; the practice round will not be paid out. These damages mean that the first word you encode will lower the total session donation from \$336 to \$335.46. The second word you encode will lower the session donation

Figure C.3: Practice round instructions part 2.

by another 54c, down to $334.92, and so on. Remember, all other people in this room can lower the total session payment with their decisions. The total damages you have caused to the charity payment in each round will be displayed at the top of your screen, below your personal earnings.

You can reduce the damages to the charity. In the practice round you will be given 3 extra letters per word. These letters are **optional**. The extra letters are displayed to the right of the initial 6 letters that you must encode:

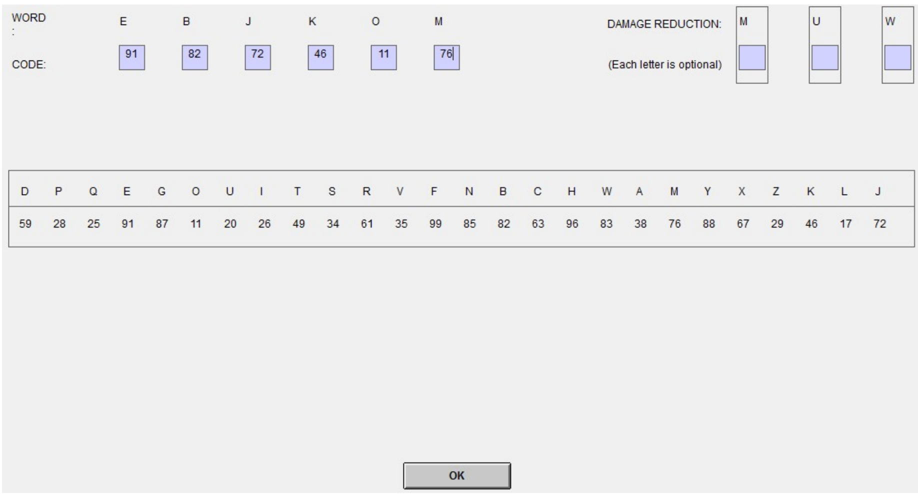


If you encode 1 of the optional extra letters, the damage to the charity for the current word you are encoding will reduce by 1 third, from 54c to 36c. You must encode this letter correctly for it to reduce the damages. If you encode 2 of the optional extra letters, the damage to the charity for the current word you are encoding will reduce by 2 thirds, from 54c to 18c. If you do not wish to cause any damage to the charity for the word you are currently encoding, you need to encode all three optional extra letters.

The money for the charity will be donated to a local environmental charity that plants indigenous trees in Victoria. Every $2 donated to the charity leads to one tree being planted. Therefore, every $2 of damages you cause to the charity's payment leads to 1 less tree being planted for that round. To prove the donation has been made, in the coming days you will receive an email with a receipt stating the total amount donated to the charity from everyone in this session. This information will include the average number of trees that will be planted per person from this session.

Examples:

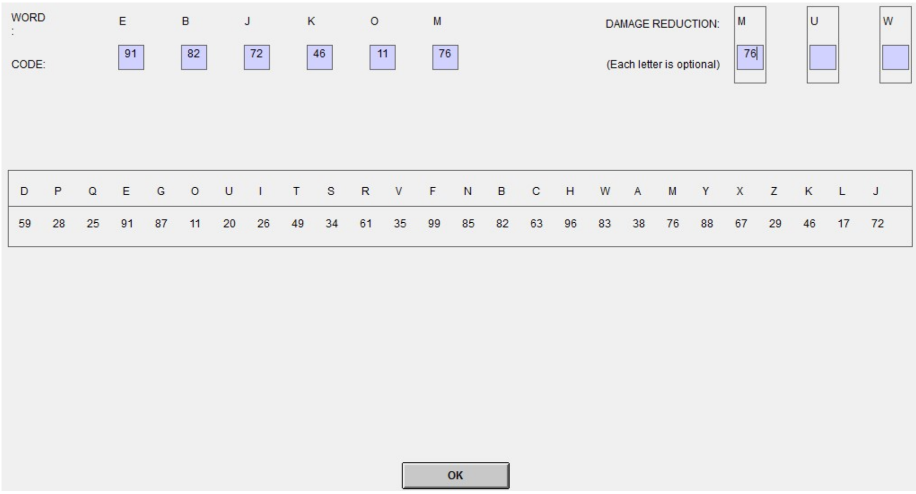
In the following image is an example of someone correctly encoding the word, and none of the optional extra letters:



They can now click OK. Once they have clicked OK, they will earn 60c and reduce the session's charity payment by 54c.

Figure C.4: Practice round instructions part 3.

In this next image, the person has added the correct code for one of the optional extra letters:



If they click OK, they will earn 60c for this word and reduce the session's charity payment by 36c.

To recap:

- You will be given a practice round. This round will not be paid out. Then you will be given 3 rounds of the task. One of these rounds will be chosen randomly by the computer to be paid out. Each round is 8 minutes (or 480 seconds) long.
- You can only use the numbers at the top of the keyboard, and you can only use your mouse to click on the boxes where you enter the numbers.
- You need to correctly encode the first 6 random letters to have a word counted. Each word earns 60c in the practice round, but given it is a practice round, these earnings will not be paid out.
- Each word you encode will cause 54c of damages to the charity in the practice round. For the word you are encoding, you can reduce the damages to 36c if you encode one of the optional extra letters, to 18c if you encode two of the optional extra letters, and to 0 if you encode three of the optional extra letters.
- Every $2 of damages you cause in the round that is paid out will lead to 1 less tree being planted.

If you have any questions about this Activity please raise your hand now.

Quiz:

You will now be given a quiz consisting of 4 quick questions to ensure you understand the activity. Feel free to refer to your printed instructions to find the correct answer.

You will earn 25c for each of the 4 questions you answer correctly. Please note that some questions have 2 parts.

The quiz will now begin.

Figure C.5: Practice round instructions part 4.

After the quiz.

Practice round:

We are about to begin the 8 minute practice round. The screen will appear exactly as it will in the actual activity. Note that the number of seconds remaining for the round is displayed at top right of the screen. Please use the practice round to familiarise yourself with the task, including how the earnings and damages are calculated.

The practice round will begin shortly.

Figure C.6: Activity instructions.

ACTIVITY Instructions

If you have any questions about the activity that have arisen from the practice round, please raise your hand now.

You will now be given the task for 3 rounds. Each round will be 8 minutes long.

Remember, just one of the 3 rounds will be randomly chosen by the computer to be paid out. Therefore, you need to treat every round as though it will be paid out.

You will only be told how much will be donated to the charity at the end of session, based on which round is chosen to be paid out.

The payment amounts for each word, and the damages per word will change between the rounds. These amounts will be displayed on your screen before the start of each round, as well as be announced aloud by the experimenter.

Are there any final questions?

The first round will begin shortly.

Figure C.7: Survey instructions.

SURVEY Instructions

You will now be given a survey. Please take your time to answer the questions truthfully.

Please note – for the second question in the survey, you will receive $1 if you answer it correctly. If you earn it, this additional $1 will be added to your final payment. The additional payment only applies to the second question. It will be clear which question the additional $1 payment applies to before you answer the question.

At the end of the survey, 1 out of the 3 rounds will be chosen to be paid out. You will be told which round will be paid out and therefore how much will be donated to the charity from this session. You will then receive your cash payment.

Are there any questions?

The survey will now begin.

Table C.5: Pro-environmental behaviours survey questions.

Now you will be asked about some environmental behaviours. Please answer honestly - remember the answers to this survey are anonymous. Thinking back over the past year, how often do you:	
1. Take shorter showers to save water	Never / Rarely / Sometimes / Often / Always
2. Turn off the tap when brushing your teeth	Never / Rarely / Sometimes / Often / Always
3. Wash only full loads of clothes	Never / Rarely / Sometimes / Often / Always /NA
4. Run the dishwasher only when full	Never / Rarely / Sometimes / Often / Always / NA
5. Fix or report leaks when you notice them	Never / Rarely / Sometimes / Often / Always / NA
6. Use the half flush button on the toilet when available	Never / Rarely / Sometimes / Often / Always
7. Put rubbish in the bin	Never / Rarely / Sometimes / Often / Always
8. Place cigarette butts in the bin	Never / Rarely / Sometimes / Often / Always / NA
9. Recycle glass, hard plastics, paper and cans	Never / Rarely / Sometimes / Often / Always
10. Use public transportation	Never / Rarely / Sometimes / Often / Always
11. Ride a bicycle for transportation	Never / Rarely / Sometimes / Often / Always
12. Choose to buy an organic option for a product if it is available	Never / Rarely / Sometimes / Often / Always
13. Choose to buy a product on the basis that it has less packaging than a similar product	Never / Rarely / Sometimes / Often / Always
14. Turn the lights off when leaving a room	Never / Rarely / Sometimes / Often / Always
15. Use re-usable bags when shopping	Never / Rarely / Sometimes / Often / Always

Table C.6: New Ecological Paradigm (NEP) survey questions (Dunlap et al., 2000).

Listed below are statements about the relationship between humans and the environment. For each one, please indicate whether you STRONGLY AGREE, MILDLY AGREE, are UNSURE, MILDLY DISAGREE or STRONGLY DISAGREE with it.
1. We are approaching the limit of the number of people the earth can support.
2. Humans have the right to modify the natural environment to suit their needs.
3. When humans interfere with nature it often produces disastrous consequences.
4. Human ingenuity will insure that we do NOT make the earth unlivable.
5. Humans are severely abusing the environment.
6. The earth has plenty of natural resources if we just learn how to develop them.
7. Plants and animals have as much right as humans to exist.
8. The balance of nature is strong enough to cope with the impacts of modern industrial nations.
9. Despite our special abilities humans are still subject to the laws of nature.
10. The so-called ecological crisis facing humankind has been greatly exaggerated.
11. The earth is like a spaceship with very limited room and resources.
12. Humans were meant to rule over the rest of nature.
13. The balance of nature is very delicate and easily upset.
14. Humans will eventually learn enough about how nature works to be able to control it.
15. If things continue on their present course, we will soon experience a major ecological catastrophe.