Putting Your Money Where Your Mouth Is

Daniel A. Brent[1,2], Lata Gangadharan[1], Anke Leroux*[1], and Paul A. Raschky[1]

[1]Department of Economics, Monash University
[2]Department of Economics, Louisiana State University

February 2016

Working Paper - please do not cite or distribute without permission

Abstract

We present a novel approach to address differences between stated and paid choices by incentivizing stated choices in a randomized field experiment. The treatment increases the incentives in the field by making each decision financially relevant. Our results show that the treatment increases the marginal utility of income, with the effect being economically and statistically significant in aggregate. The treatment also affects estimates of preferences for specific attributes by reducing willingness to pay for attributes with public benefits. Respondents with greater self-reported environmental preferences are more susceptible to the treatment in attribute space.

JEL classification: Q51, C93
Keywords: field experiment, quasi-public goods, non-market goods, stated preference, hypothetical bias

*Funding from the Cooperative Research Centre for Water Sensitive Cities (CRC grant number 20110044) is acknowledged. This project has been approved by the Monash University Human Research Ethics Committee; MUHREC#: CF12/2511 - 2912001358.

1 Introduction

Incentives are key drivers of human interactions - not only in the market place but also in diverse non-market settings, ranging from crime to marriage and family. Consistently highlighted in the works of Becker (Becker, 1993), this insight has not only influenced economic theory but also empirical research methods. For example, incentivizing individuals in economic experiments is seen as indispensable for observing behaviour in the laboratory that is consistent with behaviour in actual environments (Smith and Walker, 1993; Fellner et al., 2013). Despite the similarly heavy reliance on simulated environments, incentives have been largely missing from stated preference non-market valuation methods. Eliciting truthful values of goods and services that are non-marketed is challenging. Yet, this is precisely the area in which information about monetary values is most valuable as it helps determine the type and scope of public and quasi-public good provision. In this paper, we examine the role of incentives in the non-market valuation of projects that yield both private and public benefits.

We conduct personal interviews with a randomized sample of almost 1000 individuals to elicit their preferences for the non-market benefits of local water management. Our experimental strategy combines a discrete choice experiment with a field experiment to test the effect of providing incentives on stated preferences. A randomly determined subset of the respondents is incentivized with earned or endowed cash before choosing among alternative water management projects that vary in the provision of public and private benefits and in their costs. Prior to the choice task, respondents in the treatment group are informed that in the interest of eliciting truthful responses one of their choices will be randomly selected at the end of the experiment and the cost for the associated project will be deducted from their earnings for the benefit of a specified water management pilot project in their local community. We label this approach *salient treatment* as it connects financial incentives to the decisions of the respondents (Smith, 1982). The specific purpose of the salient treatment is to introduce a payment requirement to the survey instrument that reflects the stated choices, thereby ensuring that the coercive payment vehicle is credible.[1]

The objective of this field experiment is to evaluate two hypotheses that emerge from the mixed evidence on hypothetical bias in total and marginal willingness to pay (List et al., 2006;

[1]The randomization of the salient treatment across the survey respondents gives rise to a between-subject design, which eliminates concerns about hypothetical bias being underestimated due to respondents' desire to be consistent in their stated and revealed preferences when presented with both, hypothetical and non-hypothetical protocols (Johansson-Stenman and Svedsäter, 2008).

Johnston, 2006; Carson and Groves, 2007). First, we test if respondents are more sensitive to cost and contribute less to the quasi-public good, when faced with direct financial incentives. In the presence of hypothetical bias, one would expect to reject the null that the estimates of the marginal utility of income, the coefficient on cost from our econometric model, are the same across both treatment and control groups. Second, we test if the treatment differentially affects the preferences for the public and private benefits of the good. Provided the existing evidence on marginal willingness to pay for private goods is transferrable to quasi-public goods, one would expect to see no systematic difference in preferences across the treatment and control group for the public and private benefits of the good.

With respect to the first hypothesis, we find that the salient treatment has an economically and statistically significant effect in aggregate. In particular, the marginal utility of income increases by 85% for the treated sample as compared with the control. The treatment effects are also highly heterogeneous. The effect of the treatment on the cost parameter varies across income, with higher income groups being less responsive to the treatment than lower income groups.

We also find heterogeneous treatment effects with respect to the second hypothesis. Preferences for some benefits of the quasi-public good are statistically significantly different across the treatment and control groups. Valuation for public benefits (i.e., stream health) decreases significantly once individuals actually have to pay for their decisions. We do not find a systematic treatment effect in attributes that offer predominantly private benefits (i.e., removing water restrictions). Interestingly, we also find a large treatment effect among respondents who state a greater concern for their local environment. While we may expect that those with strong environmental preferences have a relatively high value for environmental benefits,[2] our results actually reveal that introducing financial incentives into stated choices causes the biggest decrease in valuation within this subpopulation. Overstating willingness to pay among environmentalists is critical in light of findings from Kotchen and Reiling (2000) that it is precisely those with strong environmental preferences that drive the willingness to pay for the non-marketed benefits of public environmental goods.

We discuss a number of alternative explanations for the observed differences between the control and the treatment group. These include treatment-induced incentives to answer strategically, free riding on the provision of the pilot water management project, changes to total project costs or individuals' budget constraints, as well as the effect of the initial

[2]For example, Champ et al. (1997) find a greater correlation between predicted (calibrated) and hypothetical donation behavior when respondents expressed strong attitudes about the project.

cash endowment. We conclude that these explanations are less plausible than the presence of hypothetical bias across the sampled population and especially among certain subgroups of the population.

The paper is organized as follows: Section 2 provides an overview of the relevant literature; section 3 describes the design of the field experiment and the survey; sections 4 and 5 introduce the data and present differences in average household contributions; section 6 describes the empirical framework, presents the results and discusses the treatment effects; section 7 concludes.

2 Literature Review

Our study contributes to the broad literature in economics that has defined and refined several methods of eliciting values for non-marketed goods. Throughout this process researchers have explored the advantages and disadvantages of different elicitation methods, the most commonly used method being the stated preference approach. Since the beginning of this literature, concerns about non-truthful preference revelation have dominated the debate about the reliability and external validity of stated preference surveys (Diamond and Hausman, 1994; Hanemann, 1994) and these same issues are still disputed today (Kling et al., 2012; Hausman, 2012). One prominent explanation for non-truthful preference revelation has concentrated on respondents' behavior when presented with a hypothetical scenario, while another focuses on the consequentiality and incentive compatibility of the survey instrument itself. Critics of stated preference methods term the divergence between stated and revealed choices "hypothetical bias".

Hypothetical bias could arise for two reasons. Firstly, respondents regard the implementation of the project as purely hypothetical; while secondly, they perceive that even if the project were implemented, they might not be actively involved in paying for it. These two factors can lead to biases in the estimates of economic values of non-market goods. Numerous laboratory (Harrison, 2006a,b; Andersen et al., 2006) and field experiments (Cummings et al., 1997; List et al., 2006) find evidence of hypothetical bias in the most common valuation methods with meta-analyses showing that hypothetical willingness to pay (WTP) typically exceeds the actual value by a factor of two to three (List and Gallet, 2001; Loomis, 2011). Cheap talk scripts that describe the existence of hypothetical bias have been developed (Cummings and Taylor, 1999; Ajzen et al., 2004) in an effort to mitigate this bias; however, the cheap talk fails to completely eliminate divergence between stated and revealed choices.[3] In particular, the inclusion of a cheap talk script is found to be less successful in reducing hypothetical bias in respondents that have a lot of experience with the good or are experts (List, 2001; Aadland and Caplan, 2003; Lusk, 2003). Thus, there is still an active debate about the legitimacy of stated preference methods, and potential changes in the methodology to increase their external validity.

Carson and Groves (2007) and Carson et al. (2014) oppose the concept of hypothetical bias. They argue that as long as survey questions are consequential, (e.g., the respondent believes that her answers will potentially influence whether a proposed policy or project is acted upon) and the respondent cares about the survey outcome the respondent always has incentives to answer truthfully.[4] A meta analysis by Carson et al. (1996) supports this argument and finds that WTP estimates from contingent valuation studies are highly correlated and are in the same range as household production functions or hedonic pricing studies. To address this concern our base survey design follows accepted practice to ensure the consequentiality for the respondent. Similarly, the salient treatment allows the respondents to put additional weight on their preferred choices by "putting their money where their mouth is", thereby at least weakly increasing the probability that the survey response influences local water management policy.

Moreover, it is argued that the results of many laboratory and field experimental tests of hypothetical bias are misleading as they use revelation mechanisms that are not incentive compatible in theory and/or in their practical application.[5] Therefore, results that appear to provide evidence of the existence of hypothetical bias could simply be due to the lack of incentive compatibility in the experimental design. By opting for a choice experiment with a status quo and two alternatives we purposefully selected a format that is popular for the preference elicitation for the non-market benefits of new projects or policies. Despite this design not being incentive compatible, it is the same for the control and the salient treatment and therefore is unlikely to be the reason for any observed differences between stated and paid choices.

[3] See Landry and List (2007); Özdemir et al. (2009) for examples of cheap talk scripts.

[4] The likelihood that the respondent cares about the survey outcome is enhanced if the relevant population is targeted, and the good or service to be evaluated or its attributes are plausible and are well described (Carson and Groves, 2007).

[5] For example, the theoretically incentive compatible dichotomous choice format is applied to the question of making a voluntary contribution to a public good it provides respondents with the opportunity to answer strategically: respond yes in the hypothetical treatment so as to create an opportunity for free riding in the future (Carson and Groves, 2007).

We take a neutral stance on whether past divergences in stated and revealed choices are due to hypothetical bias or lack of consequentiality in the survey instrument. However, we believe that introducing financial incentives into stated choices in a field environment is a necessary test for the validity of survey instruments. This is particularly so in settings that need to depart from the single binary choice format advocated by Carson and Groves (2007).

A recent strand of empirical studies has examined the importance of consequentiality in contingent valuation of (quasi-)public goods in settings where the conditions for achieving incentive compatibility are favorable. These typically involve the comparison of hypothetical and actual voting behavior in subsequent referenda (Vossler and Kerkvliet, 2003; Johnston, 2006; Vossler and Evans, 2009; Messer et al., 2010; Vossler and Watson, 2013) or experimental tests in the laboratory of the referendum format using real and hypothetical treatments (Cummings and Taylor, 1998; Taylor, 1998; Vossler et al., 2012). In some studies, even small non-zero probabilities of influencing the public good outcome are reported as being sufficiently consequential for a hypothetical referendum to generate outcomes that are statistically indistinguishable from the results of the actual referendum. This result is encouraging as it suggests that any positive degree of consequentiality is sufficient for truthful preference revelation in situations where the public good is well defined and its payment can be coerced.

In many field applications however, policy makers are interested in using stated preference methods as an instrument to identify the preferred scope and features of a multi-dimensional public good to be provided in the future. Aside from the difficulties in ensuring incentive compatibility of referenda in a field setting (Taylor, 1998), a distinct downside of using incentive-compatible dichotomous choice formats is that the value elicitation is insensitive to the scope of the good provided, thus making it ill suited for the estimation of marginal attribute values (Carson and Groves, 2007; Rolfe and Bennett, 2009). Multinomial choice formats (at least three choices) overcome this problem. Moreover, as incentive incompatibility primarily affects the scale factor, multinomial choice formats can be expected to provide useful estimates of the marginal trade-offs between attributes as these are independent of the scale factor (List et al., 2006; Carson and Groves, 2007). For this reason and in light of the findings on survey consequentiality, one should expect to find no differences in marginal attribute trade-offs between hypothetical and real treatments of a consequential multinomial choice survey. The experimental design in our study allows for a field examination of this argument.

3 Design of the Field Experiment and the Survey

3.1 Field Experiment

The sequence of the experiment and survey was as follows: Interviewers went to randomly selected homes, introduced themselves, and asked the householder whether he/she would be willing to participate in a survey about local water management.[6] After confirming the eligibility requirements (older than 18 years and owner-occupier status), the interviewer started the survey on an iPad. At this stage, the software randomly assigned the interviewee into the treatment ("Salient") or the control group.

The software further randomly assigned half of the treatment group into a group called "Earned Salient" and a group called "Endowed Salient". The control group immediately started with the choice task, while the "Endowed Salient" group received 1 out of 4 potential endowments (each with a probability of 0.25): A$30.60, A$39.60, A$42.00, A$53.10.[7] The "Earned Salient" group received an initial endowment of A$30.00 and, before commencing the choice task, participated in a risk elicitation task based on Holt and Laury (2002). An example of the decision problem can be found in the Appendix, Figure A.3. The earnings from this game ranged between A$0.60 and A$23.10 and were added to the respondent's initial endowment. The two salient treatments were designed such the the distribution of total earnings would a priori be comparable in both subsamples. Throughout the remainder of the survey, the respondent's money balance was shown on the upper right corner of the screen. The two salient treatments allow us to examine if project choice differs according to the source of the income obtained (endowed versus earned).[8] We do not find any differences across "Earned Salient" and "Endowed Salient", so in our analysis we pool the data for the two salient treatments.[9]

At the beginning of the choice task the interviewer carefully explained the choice situation as well as the procedure of the choice task to the respondent (see Appendix, Figure A.4). It

[6]A copy of the introduction letter can be found in the Appendix, Figure A.1. The list of households to be visited resulted from a random draw from the council's homeowner database.

[7]At the time the experiment was conducted, 1 Australian dollar was about 0.96 of the US dollar.

[8]Ideally, we would have asked participants in the treatment group to pay for the cost of their choice without first receiving a cash endowment, but the field implementation of such a design is problematic. Section 5.4 discusses potential implications of the initial cash endowment for the interpretation of our results.

[9]The test statistic of a non-parametric Mann-Whitney test for average contributions across the earned versus endowed salient treatments is 0.319 with a p-value of 0.75. In several other contexts, particularly in the laboratory, researchers have found differences in decisions between endowed and earned treatments. For examples see Cherry et al. (2002); Hoffman et al. (1994) among others.

was explicitly mentioned that we were interested in their truthful valuation of the benefits. Each individual was asked to select their preferred option out of three alternatives in 10 subsequent choice sets. Each choice set contained a status quo that was a scenario with no changes in the attribute levels with a cost of A$0, as well as two options (Options A and B) that provide improvements in at least one attribute and always had costs >A$0. These design features are constant across all respondents and treatments, so that any observed differences between the control and treatment groups can be attributed to the treatment effect. We discuss the attributes in the next subsection. In addition, the salient groups were informed that, in the interest of eliciting truthful preferences, they would be asked to randomly draw one out of their 10 choices and that the annual cost of the selected option would be substracted from their experimental earnings for the benefit a water management pilot project in their local area. The subjects' final payout was always positive and ranged between A$0.60 and A$53.10. As per information provided to the salient groups, the total amount paid by the survey participants was transferred to the respective water management pilot projects and published in the councils' newsletters.

3.2 Survey

The survey and the discrete choice experiment were designed to elicit stated preferences for urban water management in Australia. A random sample of 981 Australian individuals from four councils in Melbourne, Victoria (VIC) and Sydney, New South Wales (NSW) metropolitan areas were personally interviewed using iPads. The four councils (Fairfield [NSW], Manningham [VIC], Moonee Valley [VIC], and Warringah [NSW]) were initially chosen from a list of 29 Cooperative Research Centre (CRC) partner communities.[10] Having access to the CRC partner councils helped implement the salient treatment as these councils were inclined towards setting up pilot stormwater management projects in the near future. Similarly, residents in these councils may be more familiar with local water management initiatives, aiding the plausibility of our survey. Hence running this survey in the credible setting of a CRC partner council undertaking the proposed activity is the best attempt to provide robust, reliable, and consequential estimates even in the control group.

Among the list of partner councils, we examined several data sources to select councils that were similar along several important dimensions. First, we selected councils that were similar in the local precipitation patterns since we expect climatic factors to affect preferences

[10]The set of partner communities was established by the Cooperative Research Centre (CRC) for Water Sensitive Cities, an Australian research initiative funded by the federal government.

for water management.[11] Next, we accessed data from the HILDA database and compared the different councils along a list of demographic characteristics (income, age composition, percentage of homeowners) as well as by responses to questions about environmental preferences.[12] We selected a subset of four councils, based on their similarity in the most relevant demographic characteristics. As the objective of the study is to investigate valuation in stated preference methods, we make no claims of sample representativeness beyond seasonality, age, and home ownership status of the target population.

The survey was conducted by a professional survey company between March and August 2013.[13] Opting for personal interviews as the methodology, instead of phone, mail, or internet surveys, was important because we needed to ensure that the respondents understood the information and the alternative scenarios presented to them in the choice experiment. Moreover, the use of an iPad, with its clear visual images of the choice sets and user-friendly interface aided the respondents' understanding of the available options, thereby helping to smooth the effects of varying cognitive abilities on choices.

The goal of the survey was to inform the development of a range of different management projects for storm water, some of which were still in the planning stage without clearly specified costs and benefits. Our research thus expands on the work of Vossler et al. (2012) who ran an experiment with financial incentives for tree-planting where the project's costs and benefits were well defined. The project in Vossler et al. (2012) is relatively narrow in scope in that the key variables are the number and locations of planted trees. In contrast, the primary goal of our survey was to elicit values for several general benefits of local water management policies. This forces us to move away from the single binary choice for a well-defined project that Carson and Groves (2007) argue is most likely to achieve the conditions necessary for eliciting unbiased estimates of WTP. The lack of a well-defined project and the need to focus attention on preferences of attributes leads us to use a trichotomous discrete choice experiment. Rolfe and Bennett (2009) find that including more than two alternatives (at least two options in addition to the status quo) provides better value elicitation since a dichotomous choice masks much of the variation in specific alternatives and transforms the

[11] We accessed the daily rainfall statistics for all Australian councils from January 1890 to February 2013 from the Australian Bureau of Meteorology. We then compared long-term mean and variance in daily, weekly, and monthly precipitation between the councils.

[12] The Household, Income and Labour Dynamics in Australia (HILDA) is a government-funded Australian household panel study.

[13] This extended data collection period means that the results are not driven by the seasonality in rainfall and therefore ensures greater representativeness of the value estimates for rain-dependent attributes. Most importantly, the length of the data collection period does not adversely affect the field experiment as the selection of treatments is equally distributed over the sampling period.

decision to being primarily pro-project or anti-project. Our approach instead encourages respondents to pay attention to the attribute levels of alternatives. Therefore, this field experiment is designed to test the impact of introducing financial incentives in a complex and rich environment where valuation work is inherently challenging. As most if not all policy decisions in the field start with a scoping exercise, our approach provides important insights for the design of stated preference methods that aim to elicit direct and indirect use values of public projects. Further, and most importantly, the format and design of choice sets are constant across the treatment and control groups. Any observed treatment effects are unlikely to be due to the incentive incompatibility of the multinomial choice format.

The survey consisted of three parts: First, an introduction to the study, providing some explanation and motivation for the survey (see Appendix, Figure A.1 for the introduction letter). Second, the choice task to elicit individuals' WTP for attributes associated with local water management projects.[14] The third part of the survey was a demographic questionnaire, comprising questions on socioeconomic characteristics and attitudes towards environmental goods and services.

The attributes were selected in the following way: The CRC holds quarterly meetings of the key stakeholders involved in water management projects in Australia. These include representatives of local councils, water authorities and providers, as well as researchers from various disciplines (engineering, hydrology, climate science, urban studies, economics, law, sociology, and political science). The audience was divided into small groups, each containing at least one representative from each stakeholder group.

These small groups were asked to list the 10 most important benefits associated with storm water management. The lists were then collected and the following final set of five attributes was agreed upon in a plenary forum: Reduction in Water Restrictions, Reduction in Flash Flooding, Improvements in Stream Health, Improvements in Recreational and Amenity Benefits, and Cooler Summer Temperatures. In the next step the levels for each individual attribute were defined in collaboration with researchers from the respective disciplines. For example, the attribute levels for reduction in flash flooding were defined by a group of hydrologists, engineers, and climate scientists, while levels in the attribute Improvements in

[14] This part consisted of two choice tasks. The first one (outlined above) evaluates the benefits associated with storm water management while the second evaluates attitudes towards alternative sources of water. For this study only the first choice task is relevant. The treatment (i.e., paying for one choice set) only affected the first choice experiment. The design and sequence of the second choice task was the same for all 981 participants and its purpose was to explore preferences over alternative water sources. A copy of the second choice task is available upon request from the authors.

Stream Health were defined by a group of hydrologists, biologists, and ecologists.

In the context of water management, many attributes are subject to risk in the sense that whether or not a promised outcome is achieved also depends on exogenous factors. For example, while storm water harvesting may go some way to reduce the need for compulsory water restrictions, it may not be sufficient to achieve this outcome during a severe drought. Similarly, investment in storm water harvesting infrastructure may improve the level of biodiversity in the local stream, but the final outcome is subject to a variety of other ecological factors. In contrast, the costs of investing in storm water infrastructure are more certain. Therefore, we allow two attributes, the removal of water restrictions and improvements in stream health, to be achieved subject to some probability. We frame risk as the probability of success rather than the risk of failure. The five attributes were presented to the participants as the benefits from local water management and were defined as follows:[15]

Reductions in Water Restrictions range from a status quo scenario with no change (attribute level 1), to the exemption from less invasive restrictions (level 2), to the exemption from the most austere restrictions in the local area (level 3). The likelihood of improvement ranges from 40% probability to certainty (100% probability).

The second attribute relates to the *Reduction in Flash Flooding*. Under the status quo (level 1) the average number of flash floods over a five year period remains the same. Smaller water management projects (level 2) are able to reduce the frequency of flash floods by half, while larger water management projects (level 3) are capable of reducing the number of flash floods to almost none.

Improvements in Stream Health account for the fact that urban water management can have a direct impact on the health of local waterways. The status quo (level 1) is an unhealthy stream characterized by littered and eroded banks and low species diversity. Moderate improvements (level 2) are comprised of reduced erosion, no litter, and improved species diversity, whereas large improvements (level 3) involve the return to a diverse stream community with few nuisance species. Improvements in stream health are subject to a probability of either 40%, 60%, 80%, or 100%.

Improvements in Recreational and Amenity Benefits include, for example, recreational use benefits associated with local water ways such as paddling and swimming or the use of water for irrigation of local sports grounds and parks. The status quo (level 1) is characterized by

[15] A more detailed description of the attributes is presented in Appendix A.

rivers that are only fit to paddle, sports grounds and parks that are dry during extended periods without rain, and street line vegetation (i.e., trees) that is not watered. Moderate (level 2) recreational and amenity benefits include greener sports grounds and parks during extended dry periods and permit watering of street line vegetation. High level benefits (level 3) improve the local waterway quality to being fit for swimming and increase the amount of street line vegetation.

Cooler Summer Temperatures involve either no change in local summer temperatures under the status quo (level 1) or hot summer days being 2 degrees C cooler on average as a result of shading from additional trees being planted and evaporative cooling from artificial water bodies (level 2).

Note that the above five attributes vary in terms of their direct private benefits, with stream health having relatively more indirect benefits compared with, for example, the direct private benefits that arise from being exempt from local water restrictions.[16]

Finally, the *Costs* for the different projects are presented as additions to the household's annual water bill and range from A$0 to A$30.[17] Considering the current legal framework in Australia, this payment vehicle would also be the most likely mechanism to fund storm water management projects at the communal level. As a result, only individuals who are owner-occupiers and therefore responsible for paying the water bill, were interviewed. The selected cost levels were also chosen discussed with legal and policy experts as we were interested in realistic numbers that respondents would take credibly. The experiment was designed such that the respondents in the salient treatment would always earn more than the highest cost in any choice set. Each respondent was presented with 10 different choice sets that represented water management projects that varied along five attributes as well as costs.[18] The choice sets were generated using the NGene software package, where the D-efficiency criterion was applied to a 4x10 block design.

The questionnaire in the third part was designed to collect additional information about the respondent. A set of questions about the respondent's experience with the different attributes, the use of environmental goods, as well as experience with natural hazards is

[16]Indirect benefits are distinguished from non-use, or existence, values in the environmental economics literature and there is a rich literature on how to estimate and incorporate these values into benefit-cost analysis (see among others Lazo et al. (1997); Adamowicz et al. (1998); Common et al. (1997); McConnell (1997); Kotchen and Reiling (2000)).

[17]The explanation of the attribute "cost" reads as follows: "These are the costs per household per year of providing the water management option. These costs would be added to your annual water bill."

[18]Figure A.2 in the Appendix provides an example choice set within the explanation document.

followed by a number of questions that allow respondents to be categorized into different types depending on their attitude towards water management (i.e., their concerns for water quality and biodiversity in and around the waterways). A set of demographic controls concludes the questionnaire. Among these are an income variable, collected in intervals (based on the HILDA classification) as well as a self-reported categorization into low, medium, and high income. Both, the sequence and the content of the choice task and the questionnaire were the same across all 981 participants.

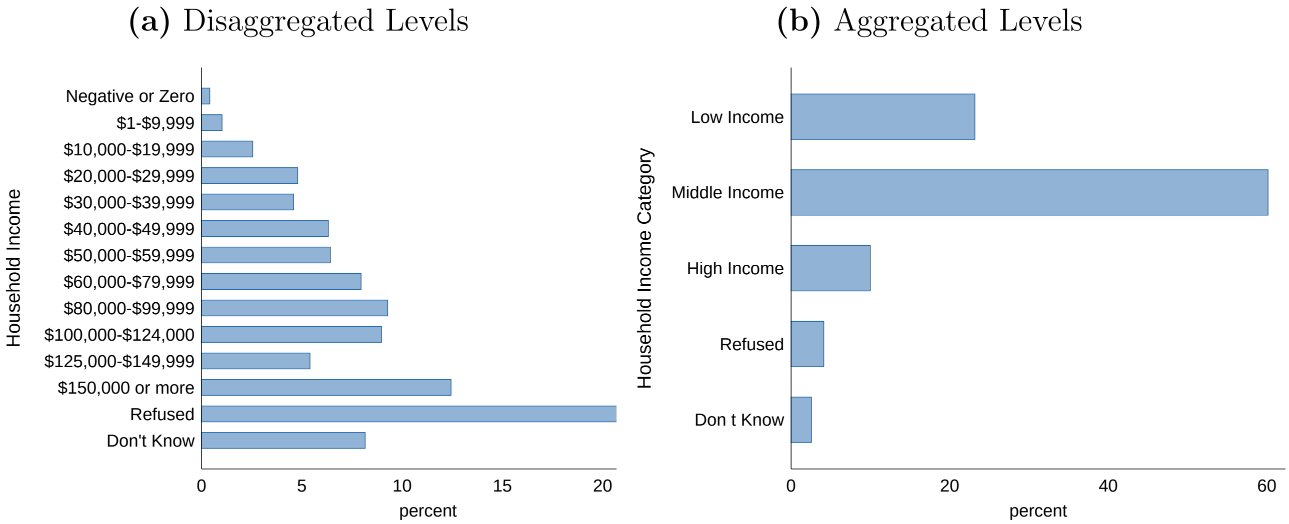
Before going into the field, interviewers were intensively briefed and trained by the authors. The fieldwork commenced with two rounds of pilot studies. The first round was conducted with a group of 10 employees from Manningham and Mooney Valley City Councils (VIC) who volunteered for the study. Of the ten volunteers, one had professional experience in local water management. The pilot was supervised by one of the authors as well as a trained social psychologist, who interviewed the volunteers before and after they completed all survey components to evaluate the overall survey design (i.e., wording, length, information content) as well as the cognitive demands of the survey. The revised version of the survey was field tested with randomly selected homeowners living in Warringah council (NSW,) before the final version was rolled out.

4 Data

We commence with the presentation of the descriptive statistics of our data set. Panel (a) of Figure 1 shows the income distribution in the sample. Many individuals refused to provide detailed information about their income, but did provide information on their general income category as seen in panel (b) of Figure 1. Since income is an important driver in our main results, we focus on the general income categories to avoid losing a considerable proportion of the data. Moreover, the respondents' perception of which income category they fall into is perhaps a more appropriate determinant of their WTP for an improvement in environmental quality. This captures their subjective income and likely incorporates their gross income relative to household expenses. Bertrand and Mullainathan (2001) find subjective data is useful at explaining behavior across individuals, as it is being used in our context.

In addition to standard demographic data, we ask questions about environmental preferences and activities that are likely to affect the willingness to contribute to a water management project. These questions include whether individuals engage in nature activities (*Nature*), if

Figure 1: Income in the Sample



they are currently facing watering restrictions (*Restrictions*), their concern for water quality (*Water Quality*), if they think a flash flood is likely or if they have experienced a flood (*Flood*) recently and whether they are concerned about increasing summer temperatures (*Summer Heat*). Some of these variables have multiple levels that we collapse into binary indicators that represent a natural division of the variable of interest.[19] This saves degrees of freedom in the estimation while still incorporating important information into the regressions. Table 1 displays the means and sample sizes of demographic and attitudinal variables for both the treatment and control group. In order to test the balance across observables, Table 1 also shows the difference in means and the p-value from a non-parametric Mann-Whitney test. The randomization procedure achieves excellent balance on observables, with only the variable Summer Heat being significantly different. An F-test for joint equality of means across treatment status cannot be rejected at the 10% level.

[19]For example, an indicator for concern over water quality takes on a value of one if the answer to the question of whether there is a need to be concerned over water quality was "very much reason" or "quite a lot of reason" and is set equal to zero if the responses are "not very much reason" or "no reason at all".

Table 1: Balance on Observables

	Mean_C	N_C	Mean_T	N_T	Difference	p-value
Low Income	0.25	603	0.25	313	-0.00	0.9849
Medium Income	0.66	600	0.63	312	0.03	0.3179
High Income	0.09	603	0.13	313	-0.03	0.1209
Age	55	647	53	332	1.75	0.1060
Female	0.46	647	0.48	334	-0.03	0.4378
Nature	0.38	647	0.36	334	0.02	0.5220
Restrictions	0.24	647	0.21	334	0.03	0.2231
Water Quality	0.37	605	0.38	315	-0.01	0.7503
Flood	0.31	630	0.33	323	-0.02	0.4976
Summer Heat	0.50	640	0.59	333	-0.09	0.0112
Joint Significance						0.1324

Notes: The columns shows the means and samples sizes for relevant demographic and attitudinal variables for both the salient group and the non-salient group, as well as the difference in means and the p-value for a non-parametric Mann-Whitney test. All variables except age are indicator variables and the means are sample proportions, and age is measured in years.

5 Nonparametric Choice Analysis

We first examine differences in the cost of selected alternatives between the treatment and control groups for subsections of the sample. Since the decisions for each respondent are likely to be highly correlated, we average the costs of selected alternatives across all choice sets for each respondent and examine differences in average costs across treatment assignment with the respondent as the unit of observation.[20] This allows us to exploit the randomization without imposing distributional assumptions on the preference parameters in a discrete choice model. Since the treatment and control groups receive the same choice sets variation in the attributes is differenced out.

The panels of Table 2 present several specifications of the average choice decisions. Each panel shows the means for the variable of interest across treatment status, samples sizes for each group, the difference in means, and the p-value from a non-parametric Mann-Whitney test.[21] Panel (a) shows the effect of treatment on the average selected cost by income level.

[20] We perform two robustness checks for the analysis of the average costs of selected alternatives. First, we use median costs instead of average costs. Second we regress the costs of selected alternatives on treatment, which captures the panel structure of the data. The regressions also control for differences in the costs of projects presented to the respondents, though the average presented project costs do not vary substantially, taking on values of A$17, A$17.75, and A$18.25 in the different choice set blocks. The results are consistent in terms of both magnitudes and statistical significance so we focus on average costs of selected alternatives.

[21] The distribution of costs is bi-modal due to a mass at zero which represents the status quo and therefore a t-test that assumes normality is not appropriate.

For the whole sample the treatment group chose projects that cost A$0.63 less than the control, representing approximately 5% of the average cost, and the Mann-Whitney test rejects equality of the two distributions at the 10% level. This shows a noticeable treatment effect even at a highly aggregated level.

We further analyze the average selected cost by respondents' income to assess heterogeneity in the treatment effect. The salient treatment primarily affects decisions for low income respondents whose average selected alternative costs A$1.74 lower compared to the control, a 14% difference that is statistically significant at the 5% level. Medium income households receiving the treatment select slightly cheaper alternatives and the high income households actually choose more expensive alternatives, although neither difference is statistically significant. Since the treatment actually induces the high income group to select more expensive projects we also restrict the sample to low and medium income; and the Mann-Whitney test rejects the null of equal distributions at the 5% level. To put the differences we observe into context, in a meta-analysis List and Gallet (2001) find that respondents overstate willingness to pay by a factor of three. Thus, relative to previous studies, the differences in chosen costs between the treatment and control groups are smaller, yet the differences in certain subgroups are statistically significant.

We envision the salient treatment working on two margins: the extensive margin represents the probability of choosing some positive payment over the status quo, which has zero cost. The intensive margin reflects the cost of a selected project conditional on paying some non-zero amount. Panel (b) shows the effect of treatment on the probability of selecting the status quo option separated by income group. The results show small differences in the probability of choosing the status quo across treatment for all income levels; the whole sample and low income subsample are significant at the 10% level while tests for the other subsamples cannot reject the null. The signs are consistent with the results for the pooling both margins as presented in panels(a).

To investigate the effect on the intensive margin, we restrict the sample to decisions where the respondent did not choose the status quo. Conditional on selecting a project that offers some improvement, we test for the impact of treatment on the *cost* of the selected alternative. The results, displayed in Table panel (c), show that all groups decrease the conditional size of the project, which is the expected result. In particular, the high income subsample, conditional on non-zero contributions, select cheaper projects if the resulting costs are immediately paid by the respondent.[22]

[22]This finding is consistent with the interviewer effect that respondents want to please the interviewer

Next we examine other demographic and attitudinal variables that we expect to drive differences in selected costs between the treatment and control groups. Panel (d) shows the differences in average selected costs by treatment across several demographic variables. Treated respondents who engage in nature and have children do not reduce their average selected costs. These groups have stronger preferences for water management as evidenced in exploratory analysis where both of these variables reduce the probability of selecting the status quo. This is consistent with the hypothesis that these groups are already predisposed to value the benefits of water management and the monetary treatment does not affect their average selections. Of all the demographic variables, the treatment has the largest difference across gender, with women reducing their average selected cost by A$1.49. This result is intriguing because on average female respondents are less likely to choose the status quo option. So while the treatment magnifies the effect of low income respondents choosing cheaper alternatives, treatment mitigates the effect of women choosing more expensive alternatives. Concern for water quality has a smaller and insignificant reduction in average selected cost for those in the treatment group.

[DB: THESE RESULTS ARE THE OPPOSITE OF THE POINT WE'RE TRYING TO MAKE WITH STREAM HEALTH. IF THE RESULT WAS ROBUST WE WOULD EXPECT TO SEE IT IN THE RAW DATA TOO. THIS IS MORE EVIDENCE FOR DROPPING THE INTRINSIC VS. EXTRINSIC MOTIVATION RESULTS, OR RETHINKING THEM. I KNOW THIS IS LOOKING AT COST AND NOT STREAM HEALTH, BUT SOME OF THE INTUITION IS THE SAME. AL: DANNY I am reluctant to drop the intrinsic vs extrinsic motivation results, I just think they are interesting even though they are now weaker. have you tried running the above analysis with concern for water quality (as per intrinsic analysis) instead of engaging in nature?]

We also test for endowment effects by comparing average selected cost in the treatment group among different levels of the initial endowment.[23] None of the differences in means are statistically different from each other and there is no monotonic relationship between the endowment and the difference in selected costs. In the next section we estimate a discrete choice model to investigate the impact of treatment on preference parameters.

---

and not appear stingy. So when respondents from high income households know they are paying with their own money they are more likely to contribute, but at lower levels. Only 9% of the sample are high income so we lose substantial statistical power for hypothesis tests of this subgroup.

[23]Table A.2 in the Appendix shows the difference in average selected cost for the salient group separated by the initial endowment.

Table 2: Cost of Selected Alternatives Across Treatment Status

(a) Income Level

	Control	N_C	Treatment	N_T	Difference	p-value
All	13.76	647	13.13	334	0.63	0.0939
Low Income	12.71	148	10.97	77	1.74	0.0411
Medium Income	14.24	395	13.90	195	0.34	0.3616
High Income	15.14	57	16.01	40	-0.87	0.7468
Low & Med Income	13.63	590	12.74	294	0.89	0.0498

(b) Income: Probability of choosing status quo

	Control	N_C	Treatment	N_T	Difference	p-value
All	0.22	647	0.25	334	-0.03	0.0924
Low Income	0.28	148	0.36	77	-0.08	0.0724
Medium Income	0.19	395	0.21	195	-0.01	0.7206
High Income	0.13	57	0.10	40	0.03	0.3137
Low & Med Income	0.23	590	0.27	294	-0.04	0.1034

(c) Income: Intensive margin contribution

	Control	N_C	Treatment	N_T	Difference	p-value
All	16.72	505	16.36	253	0.36	0.1259
Low Income	16.60	109	15.35	48	1.24	0.0826
Medium Income	16.85	316	16.61	156	0.24	0.2464
High Income	16.85	48	16.50	38	0.35	0.5535
Low & Med Income	16.70	457	16.34	215	0.37	0.1711

(d) Demographic variables

	Control	N_C	Treatment	N_T	Difference	p-value
Nature	14.80	244	14.75	119	0.05	0.8033
Children	13.83	204	13.83	121	-0.00	0.7542
Female	14.21	295	12.65	161	1.56	0.0471
Water Quality	15.74	224	15.15	120	0.58	0.1478

Notes: The columns show the average cost of selected alternatives for the Treatment group and the Control group as well as the difference in means and the p-value from a non-parametric Mann-Whitney test. The rows designate different subsections of the sample across key demographic variables.

6 Regression Framework

6.1 Econometric Model

In a random utility model (RUM) the utility function is specified to contain deterministic and random components as seen in equation 1. If the random term is additively separable then the model can be estimated by various common econometric models depending on the assumed distribution of the random component.

$$U_{ijt} = V_{ijt} + \epsilon_{ijt} \tag{1}$$

In this framework, respondent i will choose alternative j in choice t if that is the option that yields the highest level of utility. The probability of this choice occurring is displayed in equation 2.

$$\begin{aligned} \pi_{ijt} &= Pr(Y_{it} = j) = Pr(U_{ijt} > U_{iht}) : \forall : h \neq j \\ &= Pr(V_{ijt} + \epsilon_{ijt} > V_{iht} + \epsilon_{iht}) : \forall : h \neq j \end{aligned} \tag{2}$$

If we assume ϵ_{ijt} follows a type I extreme value distribution then the choice probabilities can be modeled in the logit specification shown in equation 3.

$$Pr(Y_{it} = j) = \frac{\exp(V_{ijt})}{\sum_{h \in J} \exp(V_{iht})} \tag{3}$$

In our setting, the respondents select one of three alternatives from each choice set, requiring a model that accommodates multiple categories. Based on the results of a Hausman test (Hausman and McFadden, 1984) we reject that the IIA assumption on restrictions of substitution patterns holds in our setting and therefore eliminate the standard multinomial logit as a valid econometric model. Our preferred specification is the mixed logit (MXL), which McFadden and Train (2000) show can accommodate any set of substitution patterns.[24] Additionally, the MXL model is popular in the applied literature estimating WTP from discrete choice experiments; see among others Revelt and Train (1998); Greene and Hensher (2003); Hensher et al. (2005); Balcombe et al. (2011). The MXL also allows for individual level heterogeneity by estimating a distribution of parameters across the individuals in the sample. The mixed logit has random coefficients and the probability that respondent i selects alternative j for choice t is

$$P_{ijt} = \int \frac{\exp(X'_{ijt}\beta)}{\sum_{h \in J} \exp(X'_{ijt}\beta)} f(\beta|\theta) d\beta \tag{4}$$

The choice probabilities of the MXL model therefore are weighted averages of the observable component of utility. The weights are determined by the density $f(\beta|\theta)$, where θ are the

[24] We also estimate a nested logit with the nests as the status quo and the two non-status quo options that produces similar results.

distributional statistics such as the mean and variance that are estimated from the data. There is no closed form for the parameters in the model and therefore the estimates are approximated through numerical simulation (Train, 2009).

6.2 Regression Results

The results from the base regression can be found in column (1) of Table 3. The level of each attribute is modeled as a dummy variable equal to one if that attribute-level is present for a given alternative within a choice set. We pool flood-never and flood-half as well as recreation-high and recreation-medium. The first set of parameters under the heading *Fixed Coefficients* displays the fixed coefficients followed by the means of random parameters under the heading *Random Coefficients*. All random parameters are normally distributed and the corresponding standard deviations are presented in the second section of Table 3.[25]

In order to preserve space, the standard errors and demographic variables are not presented in Table 3, rather we only display significance levels based on standard errors clustered at the respondent level. Table A.5 in the Appendix presents the full set of results with standard errors. The attributes are all modeled as random parameters and the mean of the distribution for each attribute has the expected sign with the exception of flood protection and likelihood of improvement.[26][DB: I DON'T HAVE CUMULATIVE RISKINESS IN THE REGRESSION - INCLUDING THIS REDUCES THE STATISTICAL SIGNIFICANCE OF TREATMENT ($P > 0.1$). AL: THEN MAYBE WE SHOULD NOT REFER TO IT HERE?] It is important to note the substantial heterogeneity in preferences for both the attributes and cost, as evidenced by the large standard deviations. Respondents prefer a water management alternative to the status quo, all else being equal.

Columns (2) - (4) in Table 3 include interaction terms with the treatment under the heading *Treatment Interactions (Fixed)*, which are modeled as fixed coefficients. The interaction of cost and treatment shows how the treatment impacts the respondents' sensitivity to the project cost, also interpretated as the marginal utility of income. We only present the interaction with cost, but models that replace the interaction terms with status quo yield

[25]Note that both the means and standard deviations are parameters to be estimated and therefore there are standard errors associated with both sets of parameters.

[26]One explanation why flood protection may not be desirable for respondents is that this refers to flash floods as opposed to large scale flooding. Many consumers are likely to undertake averting behavior through the purchase of private goods such as flood insurance, elevated housing, and sandbags. Thus they may not see a role for their local council in reducing the likelihood of flash floods.

similar results. Additionally, the interaction of treatment and cost captures changes on both the extensive and intensive margin because the status quo option always has a zero cost. The interaction term in column (2) is negative and both statistically and economically significant. The treatment effect is similar in magnitude to the mean of the cost coefficient, indicating that the estimated marginal utility of income increases by approximately 85% in the treatment group. Since willingness to pay is inversely scaled by the marginal utility of income, higher estimates for the marginal utility of income reduce the willingness to pay in the treated group relative to the control.

We test the heterogeneity in the treatment's impact on the estimates of marginal utility by income level to build on our results in Section 5. Similar to raw choice data we find that in magnitude the treatment effect is largest for the low income group, and the point estimate for the high income respondents is positive. While none of the treatment effects by income subgroup are statistically significant on their own, the low and high income groups' are statistically different from each other (p-value = 0.06). One reason why it is difficult to investigate treatment heterogeneity with respect to cost in a mixed logit model is that we need to include substantially more interactions with the cost attribute; consequently the standard errors on the treatment interactions with the low and high income respondents are roughly twice as large as the base treatment effect (see Table A.5).

In our setting there was no specific project to be built, and the primary focus was estimating values for the benefits that various storm water management projects could provide. The econometric analysis thus far (Columns [1] - [3] of Table 3) has implicitly assumed that the response to the treatment is independent of the attributes in the choice set and only impacts the probability and magnitude of contributions through the interactions with program cost. Our second hypothesis tests this assumption by examining if the treatment causes changes in preferences for the attributes.

We test for treatment effects within the attributes of the choice set by interacting the salient treatment dummy with each attribute. We find that most of the signs are insignificant except for medium stream health, which is significant at the 10% level.[DB: I THINK WE NEED TO REVISIT WHETHER WE WANT TO INCLUDE THIS. STREAM MED IS SIGNIFICANT AT 10% LEVEL AND STREAM HIGH IS NOT CLOSE TO SIGNIFICANT. I'M RE-RUNNING THE NATURAL VS. EXPERIMENTAL REGRESSIONS NOW.] With the exception of stream health, our results that the treatment does not have an effect on most attributes are in line with our second hypothesis and also consistent with the finding of List et al. (2006) that marginal willingness to pay for attributes is not susceptible to hypothetical bias. While List et al. (2006) only examined a private good, our setting must be viewed in the context of valuing attributes of a quasi-public good that provides different types of benefits. Lower temperatures, recreation, lifting watering restrictions, and reducing flash floods all have direct private benefits to the respondents. In contrast, stream health provides primarily indirect public benefits. The fact that most preferences for attributes are statistically indistinguishable across treatment status provides evidence that the treated respondents' focus is on the attributes of the proposed water management alternatives and not on the attributes of the pilot project.[27] The difference in preferences for stream health across treatment are explored in more detail below.

[27] Since the pilot project has fixed attributes, respondents in the treatment group may be indifferent to attributes in the choice set. If this were the case, the treatment group's preferences towards the attributes should be observationally equivalent to random behavior.

Table 3: Cost and Treatment Interactions

	(1) Base	(2) Cost	(3) Cost*Income	(4) Attributes
Fixed Coefficients & Means				
Fixed Coefficents				
Status Quo	-0.8935***	-0.9056***	-0.9986***	-0.8943***
Low Income*Cost			-0.0330***	
High Income*Cost			0.0148	
Random Coefficients				
Cost	-0.0283***	-0.0214***	-0.0178***	-0.0260***
Flood Protection (Both)	-0.1573***	-0.1498***	-0.1658***	-0.1376**
Restrictions 3,4	0.3034***	0.3087***	0.2987***	0.3022***
No Restrictions	0.2394***	0.2594***	0.2549***	0.2680***
Stream Medium	0.2400***	0.2362***	0.1986***	0.3134***
Stream High	0.2811***	0.2764***	0.2295***	0.3165***
Recreation (Both)	0.0137	0.0142	0.0192	0.0151
Temp -2	0.0858**	0.0694*	0.1060**	0.0946*
Treatment Interactions (Fixed)				
Cost*Treatment		-0.0182**		
Cost*Treatment*Low Income			-0.0226	
Cost*Treatment*Med Income			-0.0092	
Cost*Treatment*High Income			0.0182	
Flood (Both)*Treatment				-0.0420
Restrictions 3,4*Treatment				-0.0039
Restrictions None*Treatment				-0.0796
Stream medium*Treatment				-0.2317*
Stream High*Treatment				-0.1227
Recreation (Both)*Treatment				0.0395
Temp -2*Treatment				-0.0189
Std. Dev.				
Random Coefficients				
Cost	0.0980***	0.0980***	0.0968***	0.0973***
Flood Protection (Both)	-0.2695	0.2448*	0.2971**	-0.3555**
Restrictions 3,4	0.7856***	0.7622***	0.6431***	0.7947***
No Restrictions	-0.5625***	-0.5770***	0.4806***	-0.5662***
Stream Medium	0.5751***	-0.5900***	0.4749***	0.5529***
Stream High	-0.1873	-0.2474	-0.1873	0.3913***
Recreation (Both)	1.0642***	1.0624***	1.1164***	1.0580***
Temp -2	-0.5817***	-0.5678***	-0.4512***	-0.5934***
BIC	17,633.62	17,639.31	16,503.22	17,685.85
AIC	17,511.43	17,509.94	16,346.64	17,513.35
Observations	9,774	9,774	9,110	9,774
Individuals	981	981	912	981

Notes: This is a mixed logit model with random coefficients. All random coefficients are normally distributed. The heading under **Fixed Coefficients & Means** and **Std. Dev.** represent estimates of the fixed coefficients along with the means of the random coefficients and standard deviation respectively. Fixed coefficients only have an estimate of the mean. Significance levels are based on standard errors clustered at the respondent level, which are omitted to save space and are available in Table A.5. *** p<0.01, ** p<0.05, * p<0.1

6.2.1 Treatment Effects and Intrinsic Attitudinal Variables

[DB: I'M GOING TO TAKE A SHOT AT THIS SECTION, BUT WE MAY END UP CUTTING IT.] The results in Columns (4) of Table 3 indicate that the relative preferences for attributes change when respondents' actual money is tied to their choices and the good has primarily indirect public benefits. This is a particularly policy-relevant result because the relative ranking of project attributes is one of the primary motivations for eliciting values for multi-dimensional non-market public goods prior to implementing a public works project. We investigate how the treatment interacts with pre-existing variation in preferences towards the specific attributes.

Exploratory analysis on the probability of selecting the status quo option shows that respondents who are concerned about the environment are less likely to choose the status quo. This finding does not fully utilize the data because respondent characteristics are likely linked to specific project attributes. For example, while engaging in nature activities may increase the probability of choosing an alternative water management option regardless of the attributes, respondents that recently experienced water restrictions may be more inclined to select projects that lift these restrictions. We also want to test whether being predisposed to pay attention to the survey makes respondents less susceptible to the treatment.

In order to test these hypotheses we first interact attributes with variables that indicate a preference for that attribute. The combinations of attitudinal and experiential variables and the relevant attributes are presented in Table 4, and we refer to these as *intrinsic* because they are underlying characteristics of the respondents. For water quality we create an aggregate dummy that is equal to one if the respondent indicated concern for water quality or biodiversity in or around the creek, and we create a similar dummy if the respondent has had any recent experience with floods or thinks floods are likely. Since the attitudinal variables are self-reported, they are likely to be correlated with the unobserved component of the utility function that affects survey responses. We are not interpreting the results on the attitudinal variables as causal effects, but rather focus on the *difference* in the exogenous treatment assignment across certain attitudinal variables. Table 1 shows that the sample is balanced across treatment status for all the intrinsic preferences with the exception of Summer Heat. After interacting the attributes with their relevant intrinsic dummies we add an interaction with the treatment. Lastly, we interact attributes with both the intrinsic and treatment dummies.

Table 5, displays the results from the three sets of regressions as well as the base regression.

Table 4: Interactions of Intrinsic Attitudinal Variables and Attributes

Variable	Attribute
Restrictions	Restrictions 3,4 & None
Aggregate Concern for WQ	Stream Medium & Stream High
Aggregate Flood Experience	Flood (Both)
Summer Heat	Temp -2

Examining the base effect for the attributes in the *Random Coefficients* section of Table 5 we see that incorporating intrinsic preferences affects most attributes. The base effects decreases moving from column (1) to column (2) since the base effect is partly driven by respondents with stronger preferences towards those attributes as indicated by the positive and significant coefficients in the *Intrinsic Attributes* section of Table 5.

The most drastic change in parameter estimates across the specifications is for the stream health attributes. The interactions are all dummy variables and thus are interpreted as preferences for subgroups within the population. Moving from the whole population to those concerned with stream health (column [1] to column [2]) we see that the unconcerned subpopulation's coefficient is small (high stream health) or negative (medium stream health) while the concerned subpopulation's coefficients are large in magnitude and highly significant. This indicates that respondents with pro-environmental preferences are the primary drivers of aggregate preferences for stream health, which is consistent with previous work on estimating WTP for non-use values by Kotchen and Reiling (2000). However, the parameters in *Treatment Interactions* demonstrate that the estimated marginal utility from medium level stream health decreases markedly for respondents who are asked to pay for their selected choice. In column (4) of Table 5 the coefficients on the interactions between stream health, intrinsic variables, and treatment, under the heading *Both Interactions*, are negative though not statistically significant. This represents the group of respondents that care about stream health and received the treatment, which only comprises 13% of the sample and is one explanation why the coefficients are not significant. Similar to the analysis on the subsample that excludes high income households, the treatment was not randomized within the subgroups that determine the interaction variables. Our ex-post analysis shows that we achieve balance on observables for relevant subgroups.[28]

[28] Table A.4 shows that no variables are statistically significant within the subpopulation that has concern for water quality. The p-value for joint significance is 0.2087 compared to 0.1324 for the full sample.

Table 5: Attribute Interactions: Intrinsic vs. Treatment

	(1) Base	(2) Intrinsic	(3) Treatment	(4) Both
Fixed Coefficients & Means				
Fixed Coefficients				
Status Quo	-0.8935***	-0.9968***	-0.9810***	-0.9805***
Random Coefficients				
Cost	-0.0283***	-0.0286***	-0.0277***	-0.0277***
Flood Protection (Both)	-0.1573***	-0.2183***	-0.2155***	-0.1934**
Restrictions 3,4	0.3034***	0.2299***	0.2205***	0.2283**
No Restrictions	0.2394***	0.2228***	0.2620***	0.2485***
Stream Medium	0.2400***	-0.0741	0.0244	-0.0047
Stream High	0.2811***	0.0464	0.1163	0.0982
Recreation (Both)	0.0137	-0.0252	-0.0074	-0.0067
Temp -2	0.0858**	0.0615	0.0392	0.0424
Intrinsic Interactions				
Flood (Both)*Exp		0.1980**	0.2008**	0.1344
Temp -2*Concern		0.0586	0.0882	0.0844
Stream Medium*Concern		0.6935***	0.7047***	0.7870***
Stream High*Concern		0.5229***	0.4931***	0.5487***
Restrictions 3,4*Exp		0.3850***	0.3450**	0.3174*
Restrictions None*Exp		0.1959	0.1582	0.2110
Treatment Interactions				
Flood (Both)*Treatment			-0.0283	-0.0914
Restrictions 3,4*Treatment			0.1208	0.0989
Restrictions None*Treatment			-0.0618	-0.0287
Stream medium*Treatment			-0.2698**	-0.1844
Stream High*Treatment			-0.1100	-0.0546
Recreation (Both)*Treatment			0.0836	0.0820
Temp -2*Treatment			0.0135	0.0068
Both Interactions				
Flood (Both)*Exp*Treatment				0.1814
Temp -2*Concern*Treatment				0.0110
Stream Medium*Concern*Treatment				-0.2298
Stream High*Concern*Treatment				-0.1514
Restrictions 3,4*Exp*Treatment				0.0808
Restrictions None*Exp*Treatment				-0.1787
BIC	17,633.62	16,071.2	16,084.34	16,135.59
AIC	17,511.43	15,908.15	15,871.67	15,880.39
Observations	9,774	8,857	8,857	8,857
Individuals	981	888	888	888

Notes: The heading under **Means of Random Coefficients** shows the means of random parameters and **Fixed Coefficients** represents fixed parameters. Demographic variables, standard deviations of random parameters, and standard errors are suppressed. Significance levels are based on robust standard errors clustered at the respondent level. Standard deviations and standard errors are omitted to save space and are available in Table A.6 *** p<0.01, ** p<0.05, * p<0.1

The multitude of interaction terms in Table 5 make it difficult to interpret the primary effects, so we combine several parameters related to stream health in Table 6. Even though the individual coefficients are not statistically significant, the combinations of interactions produce significant comparisons. Table 6 presents the preferences for improved stream health through four subsets of the population: base, treated, intrinsic, and both intrinsic and treated. Each effect is created by a linear combination of the relevant coefficients from column (4) of Table 5. For both high and medium stream health the treatment does not have a strong effect among the population that did not indicate any concern for water quality or species. This is in contrast to those that did report concern, whom we refer to as the intrinsic group. As seen in Table 6 the preference for improving medium stream health among the concerned sample decreases by almost 50% and its statistical significance weakens when combined with the treatment. We reject the null that the concerned group has equal total preferences across treatment status at the 5% level based on the results from a chi-square test as seen in Table 6. There is a similar but slightly weaker relationship for high stream health. While we focus on the differences in preferences towards stream health across treatment, all the other attributes are relatively consistent across treatment, suggesting that preferences towards most attributes are not affected by requiring respondents to pay. These results are important because those that expressed concern for water quality and species are the primary drivers of aggregate preferences for improving stream health. In column (2) of Table 5, which includes the intrinsic interaction but does not incorporate treatment, the base effect of stream health is negative or insignificant. Our results indicate that attributes with indirect benefits are more susceptible to inflation in stated preference methods and this result is primarily driven by people who self-report greater environmental preferences

[DB: I THINK ONE OF THE PROBLEMS WITH OUR ARGUMENT NOW IS THAT THERE IS A REASONABLE LARGE (THOUGH INSIGNIFICANT) EFFECT IN THE NO CONCERN GROUP FOR MED HEALTH. THERE ARE ALSO SOME SIMILAR MAGNITUDES IN OTHER ATTRIBUTES. I LEFT THIS SECTION MAINLY THE SAME, BUT SOME OF ARGUMENTS ARE JUST A BIT WEAKER WITH THE NEW RESULTS. AL: I STILL THINK THIS RESULT IS WORTH INCLUDING - IT MAKES FOR AN INTERESTING STORY. ALTHOUGH WE HAVE TO BE MINDFUL THAT IF WE ARE THINKING OF INCLUDING THE SIMPLE THEORETICAL MODEL IT IS NOT CAPABLE OF PREDICTING THIS RESULT]

Table 6: **Intrinsic vs. Treatment for Stream Health**

	(1) Medium Stream Health	(2) High Stream Health
No Concern: Control	-0.0047	0.0982
	(0.1034)	(0.0973)
No Concern: Treatment	-0.1891	0.0436
	(0.1451)	(0.1286)
Concern: Control	0.7823***	0.6469***
	(0.1464)	(0.1336)
Concern: Treatment	0.3681*	0.4410**
	(0.1673)	(0.1644)
Observations	8,857	8,857
Individuals	888	888
No Concern: χ^2	1.17	0.13
No Concern: p-value	0.28	0.72
Concern: χ^2	3.85	1.02
Concern: p-value	0.050	0.31

Notes: The estimates are linear combinations of interactions and base effects for coefficients designated in the columns from the regression in column (4) of Table 5. The χ^2 (and accompanying p-value) are the test statistics for the hypothesis of equality for the total effect of stream health across the treatment groups in both the concerned and non-concerned subsamples. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

6.3 Discussion of Treatment Effects

A potential concern is that the differences between the control and treatment groups are not the result of increasing the salience of the payment mechanism but some other, treatment-induced effect. The first alternative explanation relates to two potential effects of the initial cash endowment on the respondents' choices. First, respondents could consider the initial endowment as additional income. Thereby, the endowment could affect the respondents' budget constraints, which can be expected to be relatively more important for low income households. However, a laxer budget constraint should lead to a greater WTP among treated lower income households and not, as we observe, a reduction in WTP. Second, respondents might consider the cash endowment as house money which they value less than their personal income. If respondents value endowed money less than earned money, we should see a difference between the "Endowed Salient" and the "Earned Salient" groups. However, there are no statistically significant differences in the contributions between those sub-groups. More importantly, a house money effect in our setting would manifest itself in the form of a lower aggregate cost sensitivity of respondents in the treatment group, while we find the opposite effect. In this context our results can be interpreted as a lower bound.

The second concern relates to project costs being greater for the treated group relative to the control. Both groups will experience increases in their annual water bills if a project is implemented, but the treatment group pays an additional upfront cost. In our experiment, the payment in the salient method is never higher than the incentives paid and also relatively minor (max. A$30) compared to the proposed payment vehicle (increase in *annual* water bills in both treatment and control), and therefore the impact of the extra payment should be negligible if the respondents consider their answers may affect council policy. Additionally, the participants in the treatment group receive benefits from their payment in the form of the increased funding for the pilot project.

Third, the treatment could affect the incentive compatibility of the survey design. For example, wealthier respondents in the treatment group might perceive that relatively poorer respondents in the treatment group are even more unlikely to opt for any positive change due to the relatively higher costs induced by the treatment. Therefore, they might act strategically and opt for a less costly option to influence the aggregate outcome. We do not find this in our data as treatment does not increase sensitivity to cost for high income respondents. However, it is difficult, if not impossible, to account for all the potential effects on strategic behavior. Finally, the treatment could be considered as capturing a donation to the pilot project and respondents in the treatment may have an incentive to understate their true willingness to pay for a good with public good characteristics. This potential to free ride on the donations of others could lead to differences in treatment and control. While this is possible, we aimed to reduce the impact of this by reminding respondents that the goal of the payment in the saliency method was to better elicit truthful answers about their value for the different benefits of the project, as opposed to raising revenue for a pilot project.

Though we cannot fully discount that some of the treatment effect may be due to free riding or strategic behavior, the existence of hypothetical bias is an important possibility to consider for policymakers who are concerned with distributional effects of projects with non-market benefits. In particular, if low income households, who generally have lower WTP, overstate their WTP, a uniform tax increase based on average WTP will exacerbate the negative welfare effects on the low income population. We observe that differences in preferences across treatment for public good benefits are most prominent among those already predisposed to pay attention and value those attributes. While List (2001) finds that cheap talk is not an effective form of reducing bias for those experienced with the good, it appears that the saliency method we are proposing is particularly effective with this group.

This suggests that including a version of the treatment may be particularly important when valuing non-market goods that cannot be valued with revealed preference approaches. This is consistent with results that hypothetical bias is greater among public goods as compared to private goods (List and Gallet, 2001), since stream health primarily provides indirect benefits that are consistent with a pure public good.

7 Conclusion

Truthful revelation is often considered the Achilles' heel of valuation methods suggested for non-market goods. This paper presents an innovative methodology to address this issue in the field by incentivizing survey respondents' choices over variants of a quasi-public good. To examine the effectiveness of this approach we conducted a discrete choice experiment to evaluate the private and public good benefits of water management projects in Australia and combined it with a field experiment. The door-to-door survey with 981 individuals took place in 2013, where a group of respondents were randomly assigned a salient treatment. The treatment group received monetary incentives prior to the actual choice task. One of their choices was then randomly selected and the cost associated with this choice was deducted from their initial monetary endowment. The money collected from the respondents was used to implement a water management project in their community. The salient treatment therefore ensures that both the payment and the provision of the good are realized, hence improving salience in both these aspects.

We find an economically and statistically significant treatment effect in aggregate. In particular the treatment manifests itself through two channels: firstly, it increases estimates of the average marginal utility of income by 85% and secondly it reduces preferences for a subset of attributes with public benefits (stream health). The reduction in preferences for stream health is primarily driven through lower values among those that self-report concern for water quality in streams. Our interpretation of these findings is that the hypothetical nature of the program cost gives rise to a bias on average; and that those predisposed towards indirect public benefits are most likely to overstate preferences in attribute space. One simple explanation of the treatment effect heterogeneity is the presence of hypothetical bias in certain subgroups. However, other channels such as free-riding and changes in strategic behavior may also contribute to differences across treatment status.

These results suggest that this saliency method can be particularly useful for stated preference studies that evaluate environmental goods with indirect benefits. Individuals who indicate environmental preferences are significantly less likely to choose projects that improve public good provision when they are required to pay for them. The elicitation of marginal willingness to pay for attributes is critical for stated preference research that aims to rank relative preferences for features of a project in the development stage. Our study also serves to address the validity of valuation methods over relevant sections of the population, which is essential when distributional effects are a concern. The saliency method can help to separate the wheat from the chaff, and improves our understanding of the true willingness to pay for such goods.

References

Aadland, David and Arthur J. Caplan, "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias," *American Journal of Agricultural Economics*, 2003, *85* (2), pp. 492–502.

Adamowicz, Wiktor, Peter Boxall, Michael Williams, and Jordan Louviere, "Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation," *American Journal of Agricultural Economics*, 1998, *80* (1), 64–75.

Ajzen, Icek, Thomas C. Brown, and Franklin Carvajal, "Explaining the Discrepancy between Intentions and Actions: The Case of Hypothetical Bias in Contingent Valuation," *Personality and Social Psychology Bulletin*, 2004, *30* (9), 1108–1121.

Andersen, Steffen, Glenn W Harrison, Morten Igel Lau, and E Elisabet Rutström, "Elicitation using multiple price list formats," *Experimental Economics*, 2006, *9* (4), 383–405.

Balcombe, Kelvin, Michael Burton, and Dan Rigby, "Skew and attribute non-attendance within the Bayesian mixed logit model," *Journal of Environmental Economics and Management*, November 2011, *62* (3), 446–461.

Becker, Gary S, "Nobel lecture: The economic way of looking at behavior," *Journal of Political Economy*, 1993, pp. 385–409.

Bertrand, Marianne and Sendhil Mullainathan, "Do people mean what they say? Implications for subjective survey data," *American Economic Review*, 2001, pp. 67–72.

Burton, M. and D. Rigby, "The Self Selection of Complexity in Choice Experiments," *American Journal of Agricultural Economics*, March 2012, *94* (3), 786–800.

Carson, Richard T and Theodore Groves, "Incentive and informational properties of preference questions," *Environmental and Resource Economics*, 2007, *37* (1), 181–210.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright, "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," *Land Economics*, 1996, *72* (1), pp. 80–99.

Carson, Richard T, Theodore Groves, and John A List, "Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice," *Journal of the Association of Environmental and Resource Economists*, 2014, *1* (1), 171–207.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum, "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods," *Journal of Environmental Economics and Management*, June 1997, *33* (2), 151–162.

Cherry, Todd L, Peter Frykblom, and Jason F Shogren, "Hardnose the dictator," *American Economic Review*, 2002, pp. 1218–1221.

Common, M, I Reid, and R Blamey, "Do existence values for cost benefit analysis exist?," *Environmental and Resource Economics*, 1997, *9* (2), 225–238.

Cummings, Ronald G and Laura O Taylor, "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method," *The American Economic Review*, 1999, *89* (3), 649–665.

Cummings, Ronald G. and Laura Osborne Taylor, "Does Realism Matter in Contingent Valuation Surveys?," *Land Economics*, 1998, *74* (2), 203–215.

_ , Steven Elliott, Glenn W. Harrison, and James J Murphy, "Are Hypothetical Referenda Incentive Compatible?," *Journal of Political Economy*, 1997, *105* (3), 609–621.

Diamond, Peter A and Jerry A Hausman, "Contingent valuation: Is some number better than no number?," *The Journal of Economic Perspectives*, 1994, *8* (4), 45–64.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler, "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information," *Journal of the European Economic Association*, 2013, *11* (3), 634–660.

Greene, William H. and David A. Hensher, "A latent class model for discrete choice analysis: contrasts with mixed logit," *Transportation Research Part B: Methodological*, September 2003, *37* (8), 681–698.

Hanemann, W Michael, "Valuing the environment through contingent valuation," *The Journal of Economic Perspectives*, 1994, *8* (4), 19–43.

Harrison, Glenn W, "Experimental evidence on alternative environmental valuation methods," *Environmental and Resource Economics*, 2006, *34* (1), 125–162.

_ , "Hypothetical bias over uncertain outcomes," *Using experimental methods in environmental and resource economics*, 2006, pp. 41–69.

Hausman, Jerry, "Contingent Valuation: From Dubious to Hopeless," *Journal of Economic Perspectives*, November 2012, *26* (4), 43–56.

_ and Daniel McFadden, "Specification Tests for the Multinomial Logit Model," *Econometrica*, 1984, *52* (5), 1219–1240.

Hensher, David A, John M Rose, and William H Greene, *Applied choice analysis: a primer*, Cambridge University Press, 2005.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith, "Preferences, property rights, and anonymity in bargaining games," *Games and Economic Behavior*, 1994, *7* (3), 346–380.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, *92* (5), pp. 1644–1655.

Johansson-Stenman, Olof and Henrik Svedsäter, "Measuring hypothetical bias in choice experiments: the importance of cognitive consistency," *The BE Journal of Economic Analysis & Policy*, 2008, *8* (1).

Johnston, Robert J., "Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum," *Journal of Environmental Economics and Management*, 2006, *52* (1), 469 – 481.

Kling, Catherine L, Daniel J Phaneuf, and Jinhua Zhao, "From Exxon to BP: Has Some Number Become Better than No Number?," *Journal of Economic Perspectives*, November 2012, *26* (4), 3–26.

Kotchen, Matthew J and Stephen D Reiling, "Environmental attitudes, motivations, and contingent valuation of nonuse values: A case study involving endangered species," *Ecological Economics*, 2000, *32* (1), 93–107.

Landry, Craig and John List, "Using ex ante approaches to obtain credible signals for value in contingent markets: Evidence from the field," Framed Field Experiments 00168, The Field Experiments Website 2007.

Lazo, Jeffrey K, Gary H McClelland, and William D Schulze, "Economic theory and psychology of non-use values," *Land Economics*, 1997, pp. 358–371.

List, John A., "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards," *American Economic Review*, 2001, *91* (5), 1498–1507.

List, John A and Craig A Gallet, "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?," *Environmental and Resource Economics*, 2001, *20*, 241–254.

_ , Paramita Sinha, and Michael H Taylor, "Using choice experiments to value nonmarket goods and services: evidence from field experiments," *Advances in Economic Analysis & Policy*, 2006, *5* (2).

Loomis, John, "What's to Know about Hypothetical Bias in Stated Preference Valuation Studies?," *Journal of Economic Surveys*, April 2011, *25* (2), 363–370.

Lusk, Jayson L., "Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice," *American Journal of Agricultural Economics*, 2003, *85* (4), pp. 840–856.

McConnell, Kenneth E, "Does altruism undermine existence value?," *Journal of Environmental Economics and Management*, 1997, *32* (1), 22–37.

McFadden, Daniel and Kenneth Train, "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 2000, *15* (5), 447–470.

Messer, Kent D., Gregory L. Poe, Daniel Rondeau, William D. Schulze, and Christian A. Vossler, "Social preferences and voting: An exploration using a novel preference revealing mechanism," *Journal of Public Economics*, 2010, *94* (34), 308 – 317.

Özdemir, Semra, F. Reed Johnson, and A. Brett Hauber, "Hypothetical bias, cheap talk, and stated willingness to pay for health care," *Journal of Health Economics*, 2009, *28* (4), 894 – 901.

Revelt, David and Kenneth Train, "Mixed logit with repeated choices: households' choices of appliance efficiency level," *Review of Economics and Statistics*, 1998, *80* (4), 647–657.

Rolfe, John and Jeff Bennett, "The impact of offering two versus three alternatives in choice modelling experiments," *Ecological Economics*, 2009, *68* (4), 1140–1148.

Smith, Vernon L, "Microeconomic Systems as an Experimental Science," *The American Economic Review*, 1982, pp. 923–955.

_ **and James M Walker**, "Monetary rewards and decision cost in experimental economics," *Economic Inquiry*, 1993, *31* (2), 245–261.

Taylor, Laura O, "Incentive compatible referenda and the valuation of environmental goods," *Agricultural and Resource Economics Review*, 1998, *27*, 132–139.

Train, Kenneth, *Discrete choice methods with simulation*, Cambridge university press, 2009.

Vossler, Christian A. and Joe Kerkvliet, "A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum," *Journal of Environmental Economics and Management*, May 2003, *45* (3), 631–649.

_ **and Mary F. Evans**, "Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality," *Journal of Environmental Economics and Management*, 2009, *58* (3), 338–345.

_ **and Sharon B. Watson**, "Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field," *Journal of Economic Behavior & Organization*, 2013, *86*, 137 – 147.

_ **, Maurice Doyon, and Daniel Rondeau**, "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments," *American Economic Journal: Microeconomics*, 2012, *4* (4), 145–71.

Appendix
For Online Publication

Definition of Attributes:

Reduction in Water Restrictions: Large parts of New South Wales and Victoria, including the urban centers Sydney and Melbourne have been experiencing persistent droughts over the last 15 years. As a result, Australian regulators have implemented water restrictions that limit the use of outdoor water. Depending on the level of the water restriction, citizens are prohibited from watering their lawn, washing their cars etc. The status quo scenario (attribute level 1) is that every level of water restriction is applicable in the local area. Some water management initiatives can mean that the respondent household and all other households in a local area will be exempt from some (attribute level 2) or all water restrictions (attribute level 3) that are imposed in the future. This attribute was also described as an attribute where the likelihood of improvement can occur with a certain degree of uncertainty (40, 60, 80, 100% likelihood that the improvements will be achieved). Exemptions from water restrictions are granted to properties in close proximity, thereby facilitating the exclusion of outsiders. Therefore, this attribute has some features of a club good.

Reduction in Flash Flooding: In the surveyed areas the major flood risk stems from pluvial or street level flooding as opposed to riverine flooding or coastal floods associated with storm surges. Pluvial flooding can occur after heavy rainfall that is not absorbed into the ground or the drainage systems due to excessive water. In urban areas, this type of street level flooding is often the result of saturated green space or an overwhelmed drainage system. Urban water management can affect the number of times street level floods (pluvial or rainfall related floods) occurring in the local area. The status quo scenario (level 1) means that the average number of flash floods over a five year period will remain the same. Smaller water management projects (level 2) are able to reduce the number of flash floods by half, while larger water management projects (level 3) are able to reduce the number of flash floods to almost none.

Improvements in Stream Health: Urban water management can have a direct impact on the health of local waterways. Healthy waterways are described to the survey participants as streams that have a diverse stream community, natural channel form and function, few nuisance species (midges, mosquitoes), and that have iconic species (platypus, frogs, native fish). The status quo scenario (level 1) was defined as a poor quality stream, with banks actively eroding, moderate to high populations of nuisance insects (mosquitoes), iconic species largely absent and litter on banks. Medium improvements (level 2) from urban water management were defined as scenarios with high quality stream community, small amounts of bank erosion, low-moderate populations of nuisance insects, some iconic species present and no litter. High improvements (level 3) can lead to situations with a diverse stream community, a natural channel form and function, low populations of nuisance insects, the presence of iconic species, no litter. Improvements in this attribute are subject to some probability of either 40, 60, 80, 100%. Arguably, compared to all other attributes in our study, improvements in stream health have more characteristics of a (local) public good and a good that has some non-use values.

Improvements in Recreational and Amenity Benefits: Urban water management can yield many recreational and amenity benefits: it influences for what activities the local waterway may be used (fit for swimming vs fit to paddle, vs not fit for contact), irrigation of local school and sports grounds during dry summers, watering of mature trees in streets and new trees planted. In the status quo scenario (level 1) the rivers are fit to paddle, sports grounds and parks are relatively dry during extended periods without rain, and street line vegetation (i.e. trees) is not watered. Medium level (level 2) recreational and amenity benefits include greener sports grounds and parks during extended dry period and permits watering of street line vegetation. High level benefits (level 3) would further make the local river fit for swimming and increase the amount of street line vegetation. Depending on the actual site (publicly accessible park vs. public sports ground with membership) this attribute has features of a public or a club good.

Costs: Costs were presented in A$5 intervals and ranged between A$0 and A$30. The upper bound of the attribute (A$30) was inferred from the costs of existing storm water management pilot projects in various partner communities. Given that the costs would be added to the household's annual water bill, this range was also approved by practitioners from local water authorities. We used a computer program (NGene) to derive the final combination of choice sets in 4 different blocks of 10 choice sets each.

Figure A.1: Introduction Letter

MONASH University

CRC for
Water Sensitive Cities

Your household has been randomly selected to take part in an **important study about water.**

The interviewer will assist you with completing the study. It will take approximately **25 minutes** to complete.

Participation is **voluntary and confidential**. Your details will not be stored with your responses and will not be passed on to any third party. You will not receive any phone calls or junk mail as a result of participating.

Details of study

Title of study

Analysis of how individuals make decisions with respect to water management in Australia.

Benefits of the study
The findings from this study will be used to help design **water management policies** in your community and Australia in general. You will also receive a monetary amount to thank you for participating, in addition to a certain amount being contributed to a water project in your local community.

What you will need to do
There are three short components to this study:
> **One and two:** Activities to examine how important the various benefits associated with different ways of managing water are to you and your community.
> **Three:** A short questionnaire.

Researchers:
This study is being conducted by I-view on behalf of Monash University researchers [redacted]

To find out more:
For more **information** on the study or to be informed of the findings after the project is complete, please contact [redacted] Monash University

☎ [redacted]
✉

If you have any **concerns** about how this study is conducted, please contact the **Executive Officer** at the Monash University Human Research Ethics Committee, quoting reference number: CF12/2511 – 2912001358:

☎03 9905 2025
✉ muhrec@monash.edu

An Australian Government Initiative

Figure A.2: Explanation Document

Explanation for Salient (without Risk) Group

ACTIVITY 1

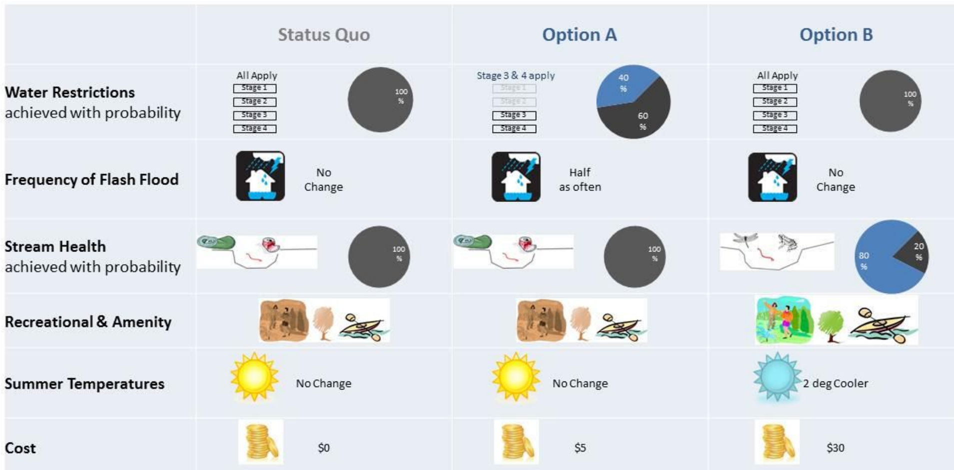
Local water management initiatives can carry a number of benefits for residents. These benefits are improvements in five key attributes, which will be explained now. Note that the improvement in two attributes, water restrictions and stream health, can be subject to uncertainty due to climatic conditions. We have therefore included pie-charts (circles) that illustrate the likelihood of a successful improvement in these attributes. The implementation success of the remaining three attributes can be considered as certain.

[USE INSTRUCTIONS CHOICE SET 1 HERE AND EXPLAIN DIFFERENT ATTRIBUTE LEVELS]

We want to understand how important these different benefits are to you. You will now be asked to make a series of 10 choices between the current situation (Status Quo) and alternative options, which involve improvements in some or all of the attributes explained above.

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

Figure A.2: Explanation Document (cont.)



- You can choose between the **Status Quo** option, **Option A** and **Option B** and you can only choose 1 option per choice set.
- The **Status Quo** option will mean:
  - No change in the current situation of water management in your council area.
  - The costs to you are zero.
- **Option A** offers **two** benefits compared with the **Status Quo**:
  - One: there is a 40% chance (as indicated by the blue area in the circle) your neighbourhood will be exempt from all future [Stage 1 and 2 [IF VIC], Level 1 and 2 [IF NSW]] water restrictions that are imposed. But, a 60% chance (as indicated by the grey area in the circle) remains that all water restrictions will apply as they do currently.
  - Two: the number of flash floods occurring in your neighbourhood will be reduced by half.
  - Choosing **Option A** would increase your annual water bill by $5. So, if this choice set were selected for payment today, $5 would be taken off your total interview earnings.
- **Option B** compared with the **Status Quo** this option
  - Carries no benefits in terms of improved water security or reduction in the frequency with which flash floods occur.
  - But, there is an 80% chance (as indicated by the blue area in the circle) that the condition of your local stream improves to medium health. A 20% chance (as indicated by the grey area in the circle) remains that there will be no improvement to local stream health compared with its current condition.
  - There are recreational and amenity benefits from keeping all local sportsgrounds and parks green and all local street trees watered during dry months.
  - Under **Option B** your local area would also be about 2 degrees Celsius cooler during the hot summer months.

Figure A.2: Explanation Document (cont.)

  - **Option B** would add $30 to your annual water bill. If this choice set was randomly selected for payment today and you had chosen **Option B**, $30 would be deducted from your interview earnings.
- Which Option would you choose? The **Status Quo**, **Option A** or **Option B?**

Your choices in this activity will help decision making on how water is managed within the community and Australia in general.

PLEASE TAKE IN TO CONSIDERATION THAT THERE ARE NO CORRECT OR WRONG DECISIONS. THESE DECISION PROBLEMS ARE NOT DESIGNED TO TEST YOU.

However, we are interested in your truthful answer about your value for these different benefits. Therefore, you should make your decisions knowing that one of the 10 choice sets will be randomly drawn by you and your final payment from this survey will be your earnings so far minus the cost of the option you have selected. Your final pay-out will always be positive but can range between $0.60 and $53.10. The full amount of money subtracted from your earnings will be donated by CRC and Monash University towards [INSERT COUNCIL WATER PROJECT], which is a project in your local area. The total amount collected from all participants will be published in [INSERT LOCAL PUBLICATION AND ISSUE DATE].

After you have completed all activities in this survey, the interviewer will ask you to randomly draw a number between **1 and 10**. This number will indicate which choice set is selected for payment and the cost of your chosen option will be deducted from your interview earnings and be put towards [INSERT COUNCIL WATER PROJECT].

In this example, **your final earnings** would have been equal to the following:

If you had chosen the Status Quo:

Your final earnings: = initial payment– $0.

If you had chosen Option A:

Your final earnings: = initial payment– $5.

If you had chosen Option B:

Your final earnings: = initial payment– $30.

Do you have any questions?

Figure A.3: Holt and Laury Lottery - Example of a Decision Problem

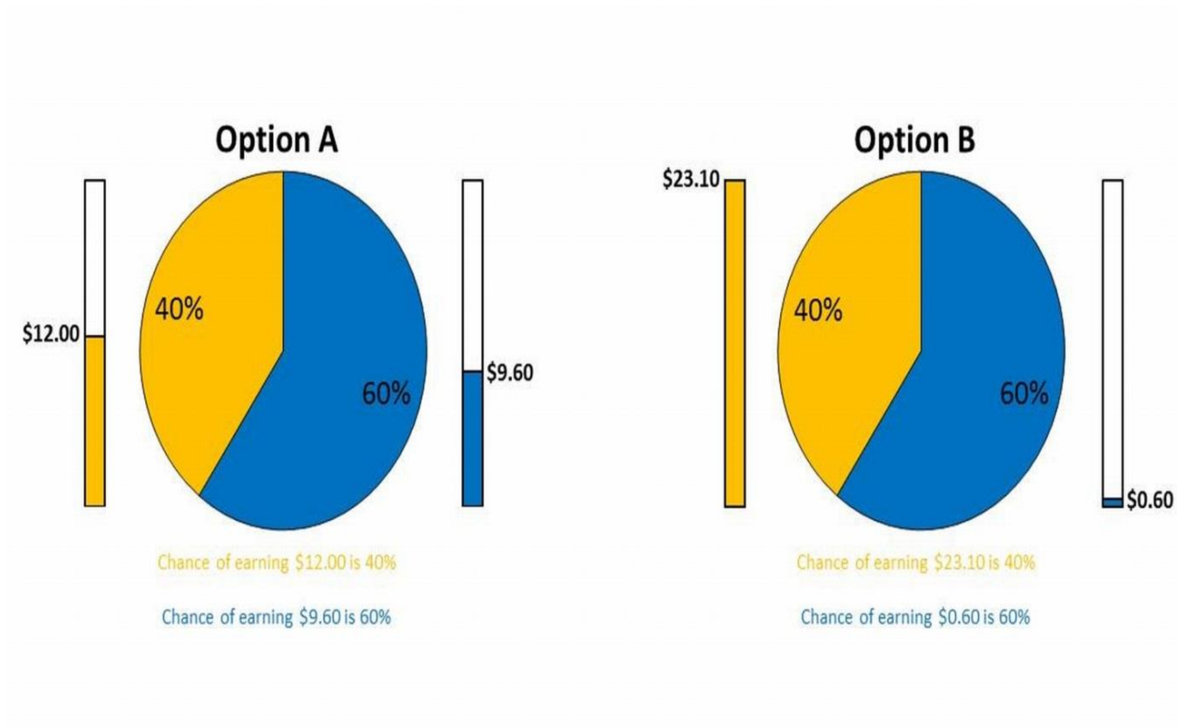


Figure A.4: Long Explanation Sheet

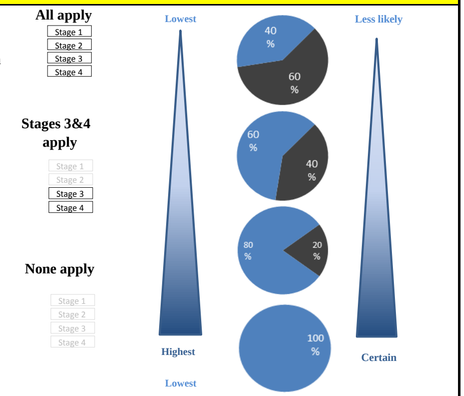
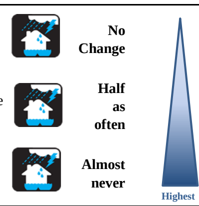
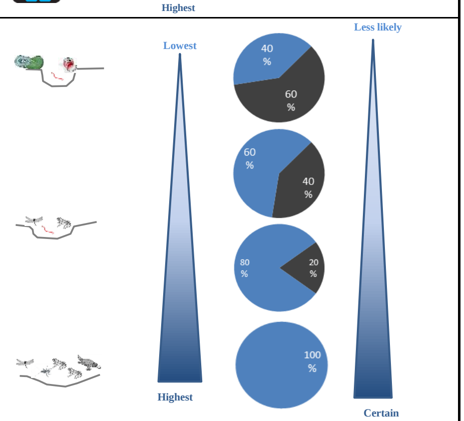
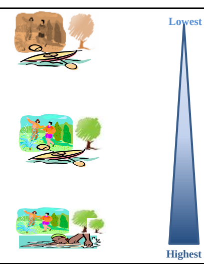
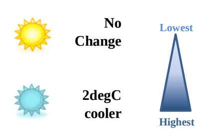
Benefit	Explanation	Levels	Visual Representation	Liklihood Improvement Occurs
Reduction in Water Restrictions	> Currently, every stage of water restriction is applicable in this local area. > Some water management initiatives can mean that your and all other households in your local area will be excempt from some or all water restrictions that are imposed in the future. This benefit is **subject to uncertainty**: > blue area in the circle illustrates how likely the improvement is > grey area shows how likely the Status Quo (no improvement) is	**Status Quo - All stages apply** > All stages of water restrictions apply in the same way as is currently the case to you and all other households in your local area. **Stages 3&4 apply** If Stage 1 or 2 water restrictions: > all households in your local area will be excempt. > watering lawns, car washing and pool filling allowed anytime If Stage 3 or 4 water restrictions: > all households in your local area need to comply as they do currently. **None apply** > No water restriction stage ever applies to your or any other household in your local area. > You could use water in the same way as if no restrictions were in place.		
Reduction in Flash Flooding	Urban water management can affect the number of times street level floods (pluvial or rainfall related floods) occur in your local area.	**Status Quo: No change** > There will be as many flash floods as there were on average in the last five years. **Half as often** > means that there will be half as many floods on average as in the last five years. **Almost never** > means that, in all likelihood, there will not be another flood in your local area.		**No uncertainty**
Improvements in Stream Health	Urban water management has direct impact on the health of your local waterway. A healty stream > has diverse stream community > natural channel form and function > few nuisance species (midges, mosquitoes) > has iconic species (platypus, frogs, native fish) This benefit is **subject to uncertainty**: > blue area in the circle illustrates how likely the improvement is > grey area shows how likely the Status Quo (no improvement) is	**Status Quo:** > poor quality stream community, > banks actively eroding, > moderate to high populations of nuisance insects (mosquitoes), > iconic species absent (platypus, frogs, native fish); > litter on banks **Medium:** > high quality stream community > small amounts of bank erosion > low-moderate populations of nuisance insects > some iconic species present > no litter **High:** > diverse stream community, > natural channel form and function, > low populations of nuisance insects, > presence of iconic species, > no litter		
Improvements in Recreational & Amenity Benefits	Urban water management can yield many recreational and amenity benefits: > it influences for what activities the local waterway may be used (fit for swimming vs fit to paddle, vs not fit for contact) > irrigation of local school and sportsgrounds during dry summers > watering of mature trees in streets and new trees planted	**Status Quo:** > river fit to paddle, > sportsgrounds and parks brown, > street trees not watered. **Medium,** > river fit to paddle, > sportsgrounds and parks green, > trees watered. **High,** river fit to swim, sportsgrounds and parks green, trees watered and new planted.		**No uncertainty**
Cooler Summer Temperatures	> Temperatures above 36 degrees Celcius cause dramatic increases in heat related discomfort and health incidents. > Urban water management has the capacity to cool urban areas by an average of 2 degrees Celcius over the summer months.	**Satus Quo: no change** > there will be no cooling in your area during summer from trees or water bodies **2degC cooler:** > your area will be on average 2 degrees cooler on hot summer days		**No uncertainty**
Cost	These are the costs per household per year of providing the water management option. These costs would be added to your annual water bill	**$0,$5,$10,$15, $20,$25, $30**	 $0, ...	**No uncertainty**

Table A.1: Difference in Probability of Choosing Status Quo

	Difference	SE
Nature	-0.12***	0.026
Restrictions	-0.10***	0.028
Water Quality	-0.13***	0.026
Flood Likely	-0.13***	0.027
Summer Heat	-0.15***	0.027

Notes: The rows represent subgroups of the population that affirmed preferences for certain environmental attributes. The difference is calculated as the proportion for the subgroup that indicated the preferences in the given row minus the subgroup without those preferences. Significance levels are from a two-sided equal proportion test and are denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Contribution between treatment: **Endowment**

	Hypothetical	N_H	Salient	N_S	Difference	p-value
Level I ($30)	13.76	647	13.17	67	0.59	0.6678
Level II ($39)	13.76	647	13.19	99	0.58	0.1443
Level III ($42)	13.76	647	13.29	84	0.47	0.5238
Level IV ($53)	13.76	647	12.88	84	0.88	0.2110

Notes: The columns show the average contribution for the salient group and the hypothetical group as well as the difference in means and the p-value for a Mann-Whitney test. The rows designate the different endowments for the salient group.

Table A.3: Balance on Observables: Exclude High Income

	Mean_C	N_C	Mean_T	N_T	Difference	p-value
Low Income	0.27	546	0.28	273	-0.01	0.7400
Medium Income	0.73	543	0.72	272	0.01	0.7514
Age	55.24	546	54.04	271	1.20	0.3418
Female	0.46	546	0.50	273	-0.04	0.2767
Nature	0.37	546	0.34	273	0.03	0.4400
Restrictions	0.24	546	0.22	273	0.02	0.5183
Water Quality	0.37	508	0.38	260	-0.01	0.6939
Flood	0.31	531	0.31	265	0.01	0.8416
Summer Heat	0.50	543	0.58	273	-0.09	0.0189
Joint Significance						0.1144

Notes: The columns shows the means and samples sizes for relevant demographic and attitudinal variables for both the salient group and the non-salient group, as well as the difference in means and the p-value for a non-parametric Mann-Whitney test. All variables except age are indicator variables and the means are sample proportions, and age is measured in years.

Table A.4: Balance on Observables: Concern for Water Quality

	Mean_C	N_C	Mean_T	N_T	Difference	p-value
Low Income	0.22	214	0.18	115	0.04	0.3765
Medium Income	0.65	213	0.69	115	-0.03	0.5298
High Income	0.12	214	0.13	115	-0.01	0.8152
Age	54.58	224	50.92	119	3.67	0.0320
Female	0.49	224	0.46	120	0.03	0.6173
Nature	0.50	224	0.44	120	0.06	0.2673
Restrictions	0.30	224	0.25	120	0.05	0.3354
Flood	0.43	218	0.42	113	0.01	0.8523
Summer Heat	0.61	221	0.69	120	-0.09	0.1183
Joint Significance						0.2087

Notes: The columns shows the means and samples sizes for relevant demographic and attitudinal variables for both the salient group and the non-salient group, as well as the difference in means and the p-value for a non-parametric Mann-Whitney test. All variables except age are indicator variables and the means are sample proportions, and age is measured in years.

Table A.5: Cost and Treatment Interactions

	(1) Base	(2) Cost	(3) Cost*Income	(4) Attributes
Fixed Coefficients & Means				
Fixed Coefficents				
Status Quo	-0.8935***	-0.9056***	-0.9986***	-0.8943***
	(0.1160)	(0.1168)	(0.1176)	(0.1166)
Low Income*Cost			-0.0330***	
			(0.0109)	
High Income*Cost			0.0148	
			(0.0133)	
Random Coefficients				
Cost	-0.0283***	-0.0214***	-0.0178***	-0.0260***
	(0.0044)	(0.0052)	(0.0067)	(0.0042)
Flood Protection (Both)	-0.1573***	-0.1498***	-0.1658***	-0.1376**
	(0.0546)	(0.0545)	(0.0553)	(0.0652)
Restrictions 3,4	0.3034***	0.3087***	0.2987***	0.3022***
	(0.0620)	(0.0611)	(0.0620)	(0.0743)
No Restrictions	0.2394***	0.2594***	0.2549***	0.2680***
	(0.0558)	(0.0564)	(0.0583)	(0.0671)
Stream Medium	0.2400***	0.2362***	0.1986***	0.3134***
	(0.0703)	(0.0718)	(0.0692)	(0.0817)
Stream High	0.2811***	0.2764***	0.2295***	0.3165***
	(0.0685)	(0.0678)	(0.0650)	(0.0766)
Recreation (Both)	0.0137	0.0142	0.0192	0.0151
	(0.0617)	(0.0603)	(0.0606)	(0.0744)
Temp -2	0.0858**	0.0694*	0.1060**	0.0946*
	(0.0402)	(0.0409)	(0.0412)	(0.0484)
Treatment Interactions (Fixed)				
Cost*Treatment		-0.0182**		
		(0.0076)		
Cost*Treatment*Low Income			-0.0226	
			(0.0175)	
Cost*Treatment*Med Income			-0.0092	
			(0.0088)	
Cost*Treatment*High Income			0.0182	
			(0.0145)	
Flood (Both)*Treatment				-0.0420
				(0.1079)
Restrictions 3,4*Treatment				-0.0039
				(0.1165)
Restrictions None*Treatment				-0.0796

				(0.1087)
Stream medium*Treatment				-0.2317*
				(0.1204)
Stream High*Treatment				-0.1227
				(0.1161)
Recreation (Both)*Treatment				0.0395
				(0.1129)
Temp -2*Treatment				-0.0189
				(0.0811)
Std. Dev.				
Random Coefficients				
Cost	0.0980***	0.0980***	0.0968***	0.0973***
	(0.0056)	(0.0058)	(0.0058)	(0.0058)
Flood Protection (Both)	-0.2695	0.2448*	0.2971**	-0.3555**
	(0.1945)	(0.1439)	(0.1332)	(0.1406)
Restrictions 3,4	0.7856***	0.7622***	0.6431***	0.7947***
	(0.0823)	(0.0848)	(0.0781)	(0.0884)
No Restrictions	-0.5625***	-0.5770***	0.4806***	-0.5662***
	(0.0880)	(0.0850)	(0.1032)	(0.0850)
Stream Medium	0.5751***	-0.5900***	0.4749***	0.5529***
	(0.1078)	(0.1024)	(0.0989)	(0.0878)
Stream High	-0.1873	-0.2474	-0.1873	0.3913***
	(0.3502)	(0.1649)	(0.2569)	(0.1240)
Recreation (Both)	1.0642***	1.0624***	1.1164***	1.0580***
	(0.0675)	(0.0616)	(0.0676)	(0.0677)
Temp -2	-0.5817***	-0.5678***	-0.4512***	-0.5934***
	(0.0641)	(0.0657)	(0.0758)	(0.0615)
BIC	17,633.62	17,639.31	16,503.22	17,685.85
AIC	17,511.43	17,509.94	16,346.64	17,513.35
Observations	9,774	9,774	9,110	9,774
Individuals	981	981	912	981

Notes: This is a mixed logit model with random coefficients. All random coefficients are normally distributed. The heading under **Fixed Coefficients & Means** and **Std. Dev.** represent estimates of the fixed coefficients along with the means of the random coefficients and standard deviation respectively. Fixed coefficients will only have an estimate of the mean. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Attribute Interactions: Natural vs. Treatment

	(1) Base	(2) Intrinsic	(3) Treatment	(4) Both
Fixed Coefficients & Means				
Fixed Coefficients				
Status Quo	-0.8935***	-0.9968***	-0.9810***	-0.9805***
	(0.1160)	(0.1220)	(0.1212)	(0.1213)
Random Coefficients				
Cost	-0.0283***	-0.0286***	-0.0277***	-0.0277***
	(0.0044)	(0.0040)	(0.0043)	(0.0044)
Flood Protection (Both)	-0.1573***	-0.2183***	-0.2155***	-0.1934**
	(0.0546)	(0.0648)	(0.0747)	(0.0773)
Restrictions 3,4	0.3034***	0.2299***	0.2205***	0.2283**
	(0.0620)	(0.0725)	(0.0851)	(0.0891)
No Restrictions	0.2394***	0.2228***	0.2620***	0.2485***
	(0.0558)	(0.0680)	(0.0789)	(0.0833)
Stream Medium	0.2400***	-0.0741	0.0244	-0.0047
	(0.0703)	(0.0855)	(0.0965)	(0.1034)
Stream High	0.2811***	0.0464	0.1163	0.0982
	(0.0685)	(0.0796)	(0.0908)	(0.0973)
Recreation (Both)	0.0137	-0.0252	-0.0074	-0.0067
	(0.0617)	(0.0625)	(0.0769)	(0.0771)
Temp -2	0.0858**	0.0615	0.0392	0.0424
	(0.0402)	(0.0631)	(0.0713)	(0.0780)
Intrinsic Interactions				
Flood (Both)*Exp		0.1980**	0.2008**	0.1344
		(0.0943)	(0.0938)	(0.1164)
Temp -2*Concern		0.0586	0.0882	0.0844
		(0.0805)	(0.0830)	(0.1027)
Stream Medium*Concern		0.6935***	0.7047***	0.7870***
		(0.1330)	(0.1384)	(0.1762)
Stream High*Concern		0.5229***	0.4931***	0.5487***
		(0.1259)	(0.1262)	(0.1622)
Restrictions 3,4*Exp		0.3850***	0.3450**	0.3174*
		(0.1355)	(0.1359)	(0.1739)
Restrictions None*Exp		0.1959	0.1582	0.2110
		(0.1281)	(0.1339)	(0.1707)
Treatment Interactions				
Flood (Both)*Treatment			-0.0283	-0.0914
			(0.1136)	(0.1330)
Restrictions 3,4*Treatment			0.1208	0.0989
			(0.1168)	(0.1346)

Restrictions None*Treatment			-0.0618	-0.0287
			(0.1110)	(0.1259)
Stream medium*Treatment			-0.2698**	-0.1844
			(0.1317)	(0.1703)
Stream High*Treatment			-0.1100	-0.0546
			(0.1216)	(0.1497)
Recreation (Both)*Treatment			0.0836	0.0820
			(0.1118)	(0.1116)
Temp -2*Treatment			0.0135	0.0068
			(0.0839)	(0.1293)
Both Interactions				
Flood (Both)*Exp*Treatment				0.1814
				(0.1969)
Temp -2*Concern*Treatment				0.0110
				(0.1700)
Stream Medium*Concern*Treatment				-0.2298
				(0.2741)
Stream High*Concern*Treatment				-0.1514
				(0.2531)
Restrictions 3,4*Exp*Treatment				0.0808
				(0.2776)
Restrictions None*Exp*Treatment				-0.1787
				(0.2742)
Std. Dev.				
Random Coefficients				
Cost	0.0980***	0.0952***	0.0934***	0.0934***
	(0.0056)	(0.0067)	(0.0061)	(0.0061)
Flood Protection (Both)	-0.2695	0.4256***	0.5586***	0.5593***
	(0.1945)	(0.1327)	(0.0968)	(0.0961)
Restrictions 3,4	0.7856***	0.6745***	0.6432***	0.6448***
	(0.0823)	(0.0953)	(0.0894)	(0.0906)
No Restrictions	-0.5625***	0.4096**	0.4199***	0.4328***
	(0.0880)	(0.1706)	(0.1536)	(0.1533)
Stream Medium	0.5751***	0.5481***	-0.6911***	-0.6916***
	(0.1078)	(0.1314)	(0.0839)	(0.0838)
Stream High	-0.1873	0.2778*	0.1908	0.1879
	(0.3502)	(0.1626)	(0.2145)	(0.2135)
Recreation (Both)	1.0642***	1.1185***	1.0917***	1.0935***
	(0.0675)	(0.0654)	(0.0648)	(0.0654)
Temp -2	-0.5817***	0.5003***	0.5367***	0.5334***
	(0.0641)	(0.0864)	(0.0737)	(0.0747)
BIC	17,633.62	16,071.2	16,084.34	16,135.59
AIC	17,511.43	15,908.15	15,871.67	15,880.39

Observations	9,774	8,857	8,857	8,857
Individuals	981	888	888	888

Notes: This is a mixed logit model with random coefficients. All random coefficients are normally distributed. The heading under **Fixed Coefficients & Means** and **Std. Dev.** represent estimates of the fixed coefficients along with the means of the random coefficients and standard deviation respectively. Fixed coefficients will only have an estimate of the mean. *** p<0.01, ** p<0.05, * p<0.1