Contingent Valuation of Organic Cotton: An Empirical Investigation into the WTA-WTP Disparity

A thesis
submitted in partial fulfilment
of the requirements for the degree
of

Master of Management Studies
at
The University of Waikato
by
Francis Powley

The University of Waikato
2011

Abstract

Contingent valuation is a non-market valuation technique which elicits preference data from participants by asking them to value a change in the provision of a specific good or service, contingent on the specifications outlined in a hypothetical market scenario. A commonly observed feature of CV results is a significant and pervasive disparity that develops between willingness to accept and willingness to pay measures of value. While these two measures should theoretically produce the same value estimate for a particular good, with the exception of a small difference due to the income effect, this is not the case in a majority of the experimental CV literature. This WTA-WTP disparity is the focus of this thesis' investigation, which aims to offer a more accurate understanding of the phenomenon.

This thesis provides a detailed review of past research experiments investigating the WTA-WTP gap, identifying two main alternate explanations for the disparity: one which explains the discrepancy as resulting from flaws in the methodological design of the CV experiments (weak experimental design) and one which suggests the gap is caused the fact that people place a higher value on a good they own than an identical good they do not own (endowment effect theory).

To assess the legitimacy of each of these two explanations, this thesis presents an experimental investigation into the WTA-WTP gap, where a basic CV survey is design and then used to elicit preference data from participants for *organic cotton*. The experimental design includes six CV surveys, all of which are fundamentally identical except for small specific alterations, which will allow valuation results to be compared across survey groups in an attempt to isolate the effect that the specified survey design features have on valuation estimates. Two of the surveys collect WTP and WTA data from participants under a binding condition where the average valuation stated by the group would determine a binding monetary outcome for all of the participants. Two further surveys collect WTP and WTA data from participants where no binding monetary outcome is specified (i.e. purely hypothetical), and the final two treatment groups are asked to estimate the WTP and WTA of the binding groups, rather than provide their own personal value estimates. The core comparisons possible between these survey designs include: assessing whether a WTA-WTP gap is observed even when controlling for features of weak experimental design, assessing how the hypothetical nature of a CV experiment impacts on the valuation results, and whether participants are able to provide an unbiased estimate of others' preferences.

Data was collected from 178 participants with between 27 and 31 respondents involved in each of the six survey groups. The data was then analysed using SPSS to test whether there were significant differences between the valuation estimates collected from the different participant groups.

The results of the experiment found that the WTA-WTP gap is caused by the endowment effect rather than weak experimental design, that hypothetical and binding treatments do not differ significantly in terms of valuation estimates, and that participants are able to provide unbiased estimates of others' preferences, so long as they are not first asked to state their own preferences.

Preface and Acknowledgements

The contingent valuation method (CVM) of non-market valuation is a new concept to me personally, and in preparing this thesis I feel I have gained a much greater understanding of this methodological process. I believe these non-market valuation techniques, such as the CVM, are going to become increasingly important in the future, as more people realise that just because something doesn't have an easily observable price, this does not mean that the good does not hold significant value. It has therefore been very enjoyable putting the contingent valuation method into practice, and I hope that the findings presented in this thesis help, in some small way, to improve the methodological design used in future CV experiments.

There are many people who I would like to thank for their contribution to the research presented within this thesis. Firstly, I must thank Professor Riccardo Scarpa, who guided me through the research process and allowed me to collect data from participants he had gathered for a separate piece of research. Without his knowledge and experience regarding non-market valuation techniques, the experimental design process would have been much more difficult.

I would of course like to acknowledge and thank all the respondents who participated in the experiment sessions. Without their contribution, the analysis of the CV method would not have been possible.

Last, but certainly not least, I would like to thank all of my family and friends for their continued support during the preparation and finalisation of this thesis. I owe them all my deepest gratitude.

Table of Contents

List of Figures

List of Tables

Glossary of Terms

Contingent Valuation	A stated preference methodology which aims to measure changes in welfare by describing a hypothetical situation to respondents and then eliciting how much they would be willing to pay either to obtain or to avoid a situation
Direct-Use value	Where individuals make actual use of the resource for either commercial purposes or recreation
Economic Value	The monetary measure of wellbeing associated with the change in the provision of some good. The term 'economic value' and 'welfare change' can be used interchangeably
Environmental Valuation	The procedure of valuing changes in an environmental good and service by measuring the changes in the producer and consumer surpluses associated with the environmental good.
Hedonic Pricing Method	A methodology which uses market price data to estimate the empirical relationship between the price of a good (e.g. housing) and the characteristics of that good (e.g. number of bedrooms, air quality, proximity to amenities, etc.). Through regression analysis, a estimated monetary value can be placed on each of the good's attributes.
Hypothetical Bias	The possibility that estimates of WTP and WTA may be biased due to the hypothetical nature of the payment commitment in stated preference surveys
Incentive Compatibility	The quality of a valuation mechanism in which truth-telling and utility maximisation coincide. An incentive compatible elicitation technique will remove any incentive for participants to employ

	strategic bargaining techniques, and ensure that answering truthfully is in the respondent's best interest.
Meta-Analysis	A statistical procedure where a number of empirical studies are used as inputs to a wider study that aims to explain the variability in the outcomes of the individual studies.
Opportunity Costs	The opportunity cost of choosing *option A* is the highest value a productive resource such as labour, capital, land or a natural resource could return if placed in the best alternative, *option B*. Opportunity cost is sometimes defined simply as 'what is given up to gain something else'.
Passive-use goods	The value placed on a resource by people who are not current users of that resource and who do not intend to use the resource themselves
Protest Bid	A response to a valuation question which does not give the respondent's genuine WTP (or WTA), but either a zero value or an unrealistically high (or low) value
Starting Point Bias	A bias present in some elicitation techniques where the final stated value of a participant shows dependence on the starting value point defined in the elicitation exercise
Strategic Bargaining	A feature of a non-incentive compatible valuation mechanism, where respondents have an incentive to either understate or overstate their true WTP or WTA in order to influence the outcome of the experiment.

Total Economic Value	Total economic value of an environmental resource is made up of i) use values and ii) non-use values. Use values are composed of a) direct use value, b) indirect use value and c) option value, whilst non-use values are made up of a) altruistic values, and b) existence values and c) bequest values
Willingness to Accept	The amount of monetary compensation an individual demands to forgo provision of a good or to accept the loss of a good.
Willingness to Pay	The amount of money an individual is willing to pay to obtain a good or to avoid the loss of a good.

List of Abbreviations

CVM	Contingent Valuation Method
RP	Revealed Preference
SP	Stated Preference
WTA	Willingness to Accept
WTP	Willingness to Pay

Chapter 1 Introduction

The contingent valuation method (CVM) is a survey-based nonmarket-valuation technique which has been used to estimate the economic value of a wide variety of goods and services, from the forests of Spain (Soliño, Prada, & Vázquez, 2010) to the opportunity to reduce one's risk of developing cancer (W. Adamowicz, Dupont, Krupnick, & Zhang, 2011). However, like all non-market valuation techniques, the CVM is not refined to a perfect science, and there is still much debate in the literature about how a CV experiment should be designed, and whether the technique is theoretically capable of estimating the true economic value of a good.

Critics of the CVM have often pointed to the significant disparity that exists between the willingness to accept (WTA) and willingness to pay (WTP) measures of value, as evidence that the methodology is unable to produce meaningful results. The WTA measure of value is collected when participants are asked to state how much compensation they would demand to give up a particular good, while the WTP measure asks participants how much they would be willing to pay to gain a good. According to standard economic theory, these two value measures should be about equal. However, significant disparities, unable to be explained by current economic theory, are persistent throughout the CV literature, and critics often cite this WTA-WTP disparity as evidence that the CVM is flawed by suggesting the disparity is caused by weak experimental design. However, proponents of the CVM argue that the WTA-WTP gap may instead indicate a fundamental upwards shift in the value that participants place on a good once they own the good.

The CVM has also been criticised in regards to the hypothetical nature of this survey technique. Some economists argue that a *hypothetical bias* would be present is CV experiments, leading to inaccurate valuation results since what people *say* they will do (i.e. how much they say they *would be* willing to pay) may not always reflect what they will *actually* do.

To assess the impact that individual design components on a CV survey have on valuation result, researchers often construct and run simple experiments

involving goods such as mugs or pens in an effort to isolate the effect of differences in survey design.

The research detailed in this thesis involves a CV experiment which has been specifically designed to assess the impact that particular survey design elements have on the valuation data. The design elements which are the primary focus of this experiment include: 1) the initial allocation of property rights for the good being valued, and the way in which the elicitation questions are phrased (i.e. WTP or WTA), and 2) the hypothetical nature of the experiment (i.e. do participant's decisions have binding monetary outcomes or are they purely hypothetical).

These survey design elements were chosen to be the focus of this thesis because there has been a great deal of experimental testing of these design elements, but disagreement still remains regarding their effects on valuation estimates.

1.1 Research Objectives

1) Investigation of the Endowment Effect

This first research objective that this thesis aims to achieve is an understanding of whether the commonly observed WTA-WTP disparity is simply the result of the methodology itself (weak experimental design) or whether the difference indicates a true difference in participants' preference towards the good depending on how the elicitation question is phrased (endowment effect). Furthermore, if the WTA-WTP disparity is indeed the result of an *endowment effect*, where participants endowed with the good place a higher value on that good simply because they own it, then this research aims to separate this effect into its two components: *parting disutility*, also known as loss aversion, and *unanticipated ownership utility*. To separate these two effects, the methodology of this research will follow an experimental design presented by Loewenstein and Adler (1995), who uses estimated WTA and actual WTA and WTP to calculate the proportion of the WTA-WTP gap caused by unanticipated ownership utility.

This area of investigation can be summarised by following two research questions:

a) Does a significant WTA-WTP disparity emerge in a CV experiment even when controlling for features of weak experimental design?
b) Are individuals who are not endowed with a good able to anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good?

2) Investigation of a Hypothetical Bias

The second feature of CV design that this thesis will address, involves investigating how the hypothetical nature of a contingent valuation experiment impacts on the valuation estimates of participants, and also to assess whether a hypothetical bias contributes to the WTA-WTP disparity. To investigate these points, two more research questions are posed:

c) Does a purely hypothetical CV experiment produce significantly different valuation results to those obtained from an identical experiment where participants' decisions have a binding monetary consequence?
d) Does a purely hypothetical CV experiment produce a significantly different WTA-WTP gap than an identical experiment with a binding monetary consequence

1.2 Thesis Overview

The structure of this thesis is as follows:

Chapter 2 will provide a detailed look at the Contingent Valuation literature focusing on investigations into the WTA-WTP disparity. This literature review will begin by introducing the concept of a non-market valuation technique, such as the CVM, explaining why these techniques are of significant important, and how they actually measure the economic value of goods without having to rely on functional market prices. Next, the process of designing a contingent valuation experiment will be looked at in detail, since the experiment this thesis is based around will adopt this methodological approach. After the CV design process is reviewed, the two measures of value which can be elicited through this methodology – WTP and WTA – will be analysed to determine that these measures are actually trying to quantify. Following this, the WTA-WTP disparity

will be introduced, and the main explanations for this disparity will be addressed: weak experimental design, the substitutability effect, and the endowment effect. Lastly, the importance of understanding the true nature of the disparity will be addressed.

Chapter 3 will describe the methodology that was used in this thesis' investigation of the four research questions detailed previously. This chapter provide an overview of the experimental design and also construct a number of hypotheses that the experiment aims to answer. Next, construction of the survey design will be described, paying particular attention to the differences between the six sets of surveys involved in the experiment. Following this, the data collection process will be reviewed, looking at how participants were collected, and how the experimental CV procedure was conducted.

Chapter 4 will provide the findings and discussion of this thesis, looking firstly at the quality and distribution of the various data sets collected during the experiment. Next, finding related to identification of the WTA-WTP gap will be presented, followed by a deconstruction of the disparity into unanticipated ownership utility and parting disutility. Following this, results will be presented to assess the impact that the hypothetical nature of the CV experiment had on the valuation results collected from participants. Lastly, biases in prediction results will be compared across the different treatment groups to determine whether these groups of participants are able to produce unbiased estimates of other participant's WTP or WTA.

Chapter 5 will provide some concluding remarks regarding the findings of this thesis, relating these findings back to the research questions detailed previously.

Chapter 2 Literature Review

2.1 Introduction

The first principal detailed in Mankiw's (2007) book *Principles of Economics [Fourth Edition]* states that "making decisions requires trading off one goal against another" (p.4). Whether it is a university student deciding how to allocate their time between socialising and studying, or a manager considering the cost of purchasing a new piece of machinery, all decisions involve a trade off. Acknowledging this trade off and having a clear understanding of what is being sacrificed is important for good decision making because "people are likely to make good decisions only if they understand the options that they have available" (Mankiw, 2007, p. 5).

In a market-based economic system, the price level of goods and services plays a crucial role in determining the appropriate allocation of resources in the economy through the concept of an *invisible hand*. Firms "decide whom to hire and what to make... [and] households decide which firms to work for and what to buy with their income", using price levels and self-interest as the guiding factors behind their choices (Mankiw, 2007, p. 9). Market economies, though based on the activities of self-interested decision makers, "have proven remarkably successful at organising economic activity in a way that promotes overall economic well-being" (Mankiw, 2007, p. 9).

However, sometimes decision making involves goods and services that do not have any kind of market price associated with them, and this makes it difficult for consumers, producers, and policymakers to make efficient decisions regarding these products. Policy makers in particular rely heavily on *cost-benefit analysis* to determine which policy design would offer the most in terms of benefits to society for the least cost (Bateman et al., 2002). If a policy decision involves non-market goods or services, such as improving the water quality of a particular lake or expanding a national park, then it is difficult for decision makers to weigh the costs against the benefits since the benefits of any decision are often intangible. With no monetary value associated the outcomes of such a policy, choosing once course of action over another, and justifying this decision, can be very difficult.

This issue of decision making and price availability was concisely expressed by Brown, Peterson, and Tonn:

> *Congressional legislation emphasises that public resource allocation should reflect the values citizens assign to those resources. Yet, information about assigned values and preferences of members of the public, including economic measures of value, required by decision makers is often incomplete or unavailable* (1995, p. 250).

One of the core goals of *environmental economics* is to assign monetary values "to non-market goods and services, where the monetary values have particular and precise meanings" (Bateman et al., 2002). Attaching such a value to these non-market goods and services allows decision makers to compare the outcomes of policy decisions in monetary terms, insuring that the most efficient decision is made. There is a significant amount of literature which investigates and critiques the various methodologies underlying this valuation process, as well as many research papers which put these techniques into practice.

This literature review begins (section 2.2) with an overview of the different approaches that can be taken to environmental-valuation broadly categorised as either revealed preference or stated preference techniques. Section 2.3 will look at the process of constructing a *contingent valuation* experiment, outlining the critical design components, and how the methodology can be tested for validity and reliability. Section 2.4 will look closely at the two different measures of value used in contingent valuation experiments, willingness to pay (WTP) and willingness to accept (WTA), and what these measures actually quantify. This will lead on to section 2.5 which introduces the concept of the WTA-WTP gap and the conventional explanation for the disparity known as the *income effect*. Following this, three alternative explanations for the WTA-WTP gap will be introduced and examined: (2.6) weak experimental design, (2.7) the substitutability hypothesis and (2.8) the endowment effect. Section 2.9 will provide justification as to why studying the true nature of the WTA-WTP gap is important for non-market valuation.

2.2 Introduction to Non-Market Valuation

This section of the literature review will explain the need for economic valuation of non-market goods and services, and will also introduce two of the most prominent methodologies for valuing such effects.

2.2.1 The Need for Non-Market Valuation Techniques

All decisions involve making a choice and all choices involve a sacrifice of some kind. If option A is chosen then option B cannot be, "if only because the resources allocated to A cannot now be allocated to B" (Bateman et al., 2002, p. 2). There are many ways that decisions can be made, and trying to assess which choice is best can be very complicated, even for decisions which at first glance may seem simple. Bateman et al. (2002) use the example of alternative ways to save people's lives to demonstrate the complexity of decision making tasks.

> Suppose there are two alternative strategies which will save human lives, both of which will cost the same amount of money to implement. Option A is expected to save 100 lives, while option B is expected to save 50 lives. Choosing option A over option B appears to be the best decision because it will save more lives. However, the question must then be asked: are all lives equal? Some may argue that saving the life of an infant if more valuable than saving the life of an elderly person, so different lives may need to be weighted to reflect this. If option A was to save the lives of 100 elderly people and option B was to save the lives of 50 infants, then choosing which of the two alternatives is best may not be as simple as previously thought. If society placed a higher value on the lives of the 50 infants, then option B may now be considered the superior strategy (Bateman et al., 2002, p. 2).

Bateman et al.'s life-saving example demonstrates how the value which society places on a particular non-market effect, in this case the age of a life to be saved, can impact on what decision should be made.

Comparing two alternative courses of action involving environmental goods, such as national parks, can also be problematic because these goods often hold recreation and intrinsic value which is not typically measured in monetary terms.

For example, suppose policy makers were choosing between option A, which would restore 100 hectares of forest, and Option B, which would restore 50 hectares of forest. Assuming both options cost the same amount of money to implement, it would appear option A offers the most value to society since it would restore twice as much forest as the alternative. However, if option B would restore 50 hectares of forest located in an area easily accessible to hikers and campers, while option A would restore 100 hectares of forest inaccessible to such recreational users, then option B might now be considered to offer the most value to society.

In both of the above example, deciding which of the two options offers the greatest benefit to society should not be based solely on how many lives are saved or how much forest is restored, but the value that society would place on those changes. However, these non-market effects often do not have any measure of value associated with them, and are therefore difficult to incorporate into decision making tasks. Environmental valuation aims to fill this information gap by measuring the economic value that society places on these non-market goods and services. This allows decision makers to weigh the true costs and benefits of alternative strategies more accurately, resulting in a more efficient allocation of resources and greater benefits to society (Bateman et al., 2002).

The terms *non-market valuation* and *environmental valuation* are closely related, but are not identical concepts. Non-market valuation is the process of estimating an economic value for a good (or an attribute of a good) based on the preferences of society in a way that does not require the presence of tradition market functions (e.g. buying and selling). Environmental valuation on the other hand, is the non-market valuation of environmental goods or services. Despite being slightly different concepts, the two terms are sometimes used interchangeable throughout the following literature review.

2.2.2 What is Non-Market Valuation?

According to Lipton, Wellman, Sheifer, Weiher, and NOAA "environmental valuation is a series of techniques that economists use to assess the economic value of market and non-market goods" (1995, p. 16). The "term *value* in economics has a precise definition — it is the price individuals are willing to pay

in order to obtain a good or service", not to be confused with *monetary value* which is the price individuals are actually required to pay (Lipton et al., 1995, p. 9). A "fundamental distinction between the way economics and other disciplines such as ecology use the term value is the economic emphasis on human preferences" (Lipton et al., 1995, p. 10). If people prefer one item over another, then that item is considered to be more valuable. By analysing the preferences of an individual towards a non-market good, or even individual attributes of a good, it is possible to estimate their willingness to pay (WTP), and therefore the value that they place on the good.

Through econometric analysis of consumer preference data collected using the methods outlined below, it is possible to estimate the economic value (i.e. their WTP) that society places on non-market goods and services such as outdoor recreation areas, wildlife, improvements in health, and many more. The ability to attach an economic value to a good or service which does not automatically have such a value assigned to it, can allow politicians to efficiently allocate capital into projects that generate the most welfare to society, and allow market research companies to design products that best cater to the demands of the consumer (Merino, 2003).

2.2.3 Revealed Preference and Stated Preference Techniques

Consumer preferences can be elicited using either Revealed Preference (RP) or Stated Preference (SP) techniques (Merino, 2003). The revealed preference approach uses existing "behavioural data to estimate the *ex post* willingness to pay [of consumers] for various commodities", and includes techniques such as the *travel cost* method and the *hedonic pricing* method (Whitehead, Pattanayak, Van Houtven, & Gelso, 2008, p. 873). Revealed preference methodologies rely on observations of the past behaviour of individuals and are therefore limited to valuing goods for which some kind of use or price data is already available.

The stated preference approach however, uses "hypothetical data to estimate *ex ante* willingness to pay for various commodities", and includes the *contingent valuation* method the *choice experiment* method, and the *contingent ranking* method, among others (Whitehead et al., 2008, p. 873). These stated preference techniques directly measures the preferences of participants by asking them

directly how much they would be willing to pay for the provision or a good (contingent valuation), or by asking them to choose their most preferred option (choice experiment) or rank (contingent ranking) carefully crafted alternative product options which range in attribute qualities and price (Bateman et al., 2002, p. 21). These SP techniques do not require any existing behavioural data from individuals, and can therefore be utilised for a wide variety of applications, such as assessing demand for products which do not yet exist (Carson, 1989).

To compare how revealed preference and stated preference techniques differ on a fundamental level, the travel cost method (RP) and the contingent valuation method (SP) will now be introduced.

2.2.3.1 ***The Travel Cost Method***

The *travel cost* method (TCM) is mostly used "for estimating the benefits of access to natural areas for tourism and recreation" such as lakes and national parks (Driml, 2002, p. 12). Since most natural recreation areas do not charge an access fee (and if they do the fee is often minimal) no traditional market exists where the benefits of the resource can be valued in terms of the price that people would be willing pay to access them. The TCM uses the cost of travelling to a particular recreation area, and the time-costs associated with the trip, as a substitute for the *price* that people pay to gain the benefits of the site (Lipton et al., 1995).

Researchers adopting the travel cost methodology must collect data from individuals on the number of trips they takes to various recreation sites each year, the distance they travel to get there, as well as their primary motivation for going to each site (e.g. tramping, relaxation, hunting and fishing, ect). By "observing the characteristics of individuals visiting the site - for example, the specific attributes of their trip to and from the site as well as the total number of visits - economists are able to estimate the 'derived demand' for the site" (Lipton et al., 1995, p. 43). The researcher then needs to isolate the attributes of the recreation site that they wish to value, and measure the level of each attribute at each recreation area. These attributes may be the quality of scenery, the amount of wildlife in the area, or any other attribute which offers some kind of benefit to a person visiting the site.

Once the appropriate data is collected, the researcher can then compare the demand function for the different recreation areas against the varying levels of benefit attained at each location. Using a random utility model, willingness to pay estimates can then be derived for the different site attributes (Shaw, 2005). See Anderson (2010) for an demonstration of the TCM being used to estimate the demand for ice climbing in Hyalite Canyon, Montana, or Bestard and Font (2009) who use the TCM to assess how environmental diversity influences which forests people in Spain choose to visit.

2.2.3.2 Contingent Valuation Method

The contingent valuation method (CVM) "is a survey or questionnaire-based approach to the valuation of non-market goods and services... [which derives] values through the elicitation of respondents' willingness-to-pay to prevent injuries to natural resources or to restore injured natural resources", though the method is not limited to just environmental goods (Lipton et al., 1995, p. 50). The CVM is a stated preference technique since it directly asks respondents how much they would be willing to pay to gain or improve a particular good, and therefore does not rely or require any existing behavioural data.

The contingent valuation method (CVM) elicits preference data directly from individuals who are placed within a carefully designed hypothetical market scenario (Mitchell & Carson, 1989). The scenario describes the good or service being valued, defines an initial level of provision, a proposed change in provision, and the conditions under which the change would provided or withheld. The respondents are then asked how much they would be willing to pay for the increase (WTP), or how much they would need to be paid in compensation to accept a decrease (WTA) in the level of the good. There are other measures of value which can be elicited using the CVM, but these will be examined in greater detail later in the literature review. This WTP or WTA data can then be used to derive the total economic gain (or loss) of the proposed change under analysis. See Du Preez, Tessendorf, and Hosking (2010) for an application of the CVM being used to value a water quality restoration project in South Africa.

2.2.4 Critique of the Revealed Preference Approach

Revealed preference valuation techniques have several shortfalls which stated preference methods, such as *contingent valuation*, aim to avoid. Firstly, "revealed-preference methods involve econometric problems that have yet to be fully overcome... The travel cost method, for example, encounters as yet unresolved problems associated with how to value time spent in travel and at the recreation site, how to treat trips with several destinations..., and other such problems" (Bishop & Romano, 1998, p. 5).

Bishop and Romano (1998) also suggest that even if these econometric issues were solved, the fundamental process involved in collected revealed preference data would still provide only a partial measure of the true value of a non-market good (p.5). For example, it may be possible to capture the increase in recreation value of a lake after an improvement in its water quality through the travel cost method. However, this would not include the increased aesthetic value observed by people driving past the lake on their way to work each morning. Furthermore, "revealed-preference methods are particularly limited in addressing 'non-use' or 'passive use' values" which originate from human preference for nature, unrelated to personal use of the environmental good itself (Bishop & Romano, 1998, p. 5).

Another limitation of RP techniques is that they rely on observations of how people behave in relation to the good being valued (e.g. their travelling behaviour utilised in the TCM), and are thus limited to valuing goods for which this information is already available. Revealed preference techniques are therefore unable to measure the expected economic value of newly developed goods, services, or unique environmental restoration policies.

The major benefit of RP methodologies, which is also its main shortfall, is that these techniques use actual behaviour rather than stated behaviour as the foundation for their valuation (as in SP techniques), and some economists argue that actual observed behaviour is the superior indicator of people's true preferences (Kahn & Tice, 1973). consider that "perhaps we can most easily glean the truth in what we see people actually doing, not from listening to what they say they will do (p. 183). However other authors, such as Wardman (1988),

have found evidence to suggest stated behaviour is a "reasonably accurate guide to true underlying preferences" (p.71).

2.2.5 Critique of the Stated Preference Approach

The main criticism of the SP approach is that "individuals' stated preferences may not correspond closely to their actual preference... because of systematic bias in SP responses... or because of difficulty in carrying out the SP task" (Wardman, 1988, p. 71). This criticism is based on the fact that RP techniques elicit preferences from participants in a *hypothetical* market environment, often using valuation mechanisms which are unfamiliar and potentially confusing to respondents. The hypothetical nature of all stated preference techniques, along with specific issues related to the various methodologies within this framework may create a systematic bias in respondent's answers so that the valuation does not reflect their true preferences.

For example, in a hypothetical contingent valuation scenario participants may be asked how much they would be willing to pay for a specific improvement in an environmental amenity. Since the participants are aware that the payment is purely hypothetical, they may be expected to overstate their true WTP for the change in an attempt to appear more generous (Bryan & Jowett, 2009), Furthermore, it is possible that participants may simply not put enough thought into their decisions, since they feel their choices do not have any real impact or meaning. Therefore, the hypothetical nature of SP techniques may need to inaccurate estimates of respondents' true WTP.

The second point which was raised by Wardman (1988) is that participants involved in SP experiments may find it difficult to provide an accurate representation of their true preference because of the unfamiliar and potentially confusing nature of the experimental market that they are placed in. Contingent valuation is more prone to this kind of error than other SP techniques such as choice experiments, especially if the scenario is considered unrealistic or confusing (see section 2.6). If the hypothetical market environment described in a CV experiment lacks realism, then participants may find it difficult to reveal their true preferences (Plott & Zeiler, 2005).

Surveys used for SP experiments are also vulnerable to response effects, where small changes in the wording of questions or the ordering or choices can sometimes cause significant changes in survey responses (Schuman & Presse, 1981). In contingent valuation for example, a response effect known as the *starting point bias* is sometimes observed when a bidding-game elicitation mechanisms is used, which creates a correlation between the level of the first bid amount and the final bid amount (Onwujekwe & Nwagbo, 2002).

The main benefit of stated preference techniques is that the "models are capable of measuring a full range of values, including so called passive use or non-use values, as they do not rely on the observation of actual behaviour" (Du Preez et al., 2010, pp. 136-137). Despite the many limitations of the stated preference approach to environmental valuation, the methodology has become increasingly popular due to its wide range of applications, and many economists and behavioural philologists are working to refine and improve the various techniques (Bateman et al., 2002).

In the following section the contingent valuation method (CVM) of non-market valuation will be analysed in further detail, providing an overview of the process involved in designing a CV experiment, and methods for testing the reliability and validity of valuation estimates.

2.3 **Designing a Contingent Valuation Experiment**

Placing a monetary value on the cost of pollution, the aesthetic and recreational value of a national park, or improvement in the water quality of a lake is a cornerstone of the economic approach to the environment (Hanemann, 1994). However, functional markets rarely exist for these types of environmental goods, so researchers have developed various methods of non-market valuation, most of which are based on the preferences of the population. Contingent valuation (CV) is one such method, categorised as a *stated preference* technique, which aims value non-market goods by collecting preference data from individuals through their response to valuation questions which are contingent on a particular hypothetical scenario. At the heart of any CV survey is the 'scenario' which describes "the attributes of the good or service to be valued and specifies the conditions under which respondents are to decide how much, if anything, they would be willing to pay for the good or service described" (Bishop & Romano, 1998, pp. 3-4). Contingent valuation "circumvents the absence of markets for public goods by presenting consumers with hypothetical markets in which they have the opportunity to buy [or sell] the good in question" (Carson, 1989, pp. 2-3).

This section will begin with a brief overview of contingent valuation, followed by an examination of the important design components involved in a contingent valuation experiment, which include: (1) designing the hypothetical market scenario, (2) choosing an elicitation mechanism, and (3) collecting characteristics data from participants. Lastly, the techniques used to assess the validity and reliability of valuation estimates elicited through the CVM will be introduced.

It is important to note that this review will not provide a comprehensive examination of the CV process in its entirety, but will rather focus on the particular design aspects important in constructing a CV experiment. Since the focus of this thesis is on the WTA-WTP gap rather than the process of calculating welfare gains or losses, reviewing the entire valuation process from experimental design to data-analysis would be largely irrelevant. For a full understanding of the process involved in defining a CV experiment, and measuring welfare changes

using CV data, see Carson, 1989; Bishop and Romano, 1998; Bateman et al., 2002.

2.3.1 Overview of the Contingent Valuation Method

Though resource-economist Ciriacy-Wantrup suggested using a *direct interview method* for valuing nonmarket natural resources as early as 1947, the first clear example of contingent valuation was used by Robert K. Davis in the 1960s when he used questionnaires to estimate the benefits of outdoor recreation in a backwoods area of Maine (Mitchell & Carson, 1989). The CVM has been used by many researchers over the past few decades to place monetary values on goods which cannot be valued through traditional market functions (i.e. price levels).

Bateman et al. (2002) explain the nature of economic valuation through the contingent valuation method (CVM) in terms of measuring the *benefits* and *costs* associated with the gain or loss of a particular non-market good. These benefits and costs are defined by an individual's personal preference towards the non-market good and their unique set of indifference curves.

> *An individual receives a benefit whenever he receives something in return for which he is willing to giving up something else that he values..., [while] an individual incurs a cost whenever she gives something that she would willingly give up only if she was given something else that she valued as compensation.*
>
> (Bateman et al., 2002, p. 16)

To determine how large the *benefit* is that an individual gains from receiving a good or service, Bateman et al. recommends measuring how much he or she is willing to pay to receive that good or service, known as their *willingness to pay* or WTP (2002, p.16). To determine the *cost* of giving up a good or service, it is suggested to measure the monetary compensation that an individual demands so that they are indifferent between retaining the good or service and the receiving the compensation rate. The level of compensation required by the individual is known as their *willingness to accept*, or WTA. Further detail regarding these two different measures of value can be found in section 2.4.1.

Estimating the true maximum amount of money a person is willing to pay to receive a good, gives an indication of the total *benefit* that person expects to gain from the good in terms of changes in their personal utility (Bateman et al., 2002). Collecting valuation estimates from a representative sample of a population, assuming the data is a true reflection of the sample's preference towards the good, provides an measure of the *total benefit* to the population that provision of the non-market good would create.

Because the CVM elicits valuation data from participants within a hypothetical market, rather than an authentic functional market, the design of the hypothetical scenario and the processes used to collect the valuation data are critical. Unrealistic or confusing market structures, or elicitation questions which create a bias in participant's responses, will produce valuation estimates which do not reflect the true value people place on the non-market good. Any conclusions based on biased WTA or WTP estimates will be inaccurate

The closer the contingent valuation experiment mimics real market processes, the closer the respondent's stated WTP should be to their true WTP (Carson, 1989). What follows is a description of how researchers design a contingent valuation experiment, with the intention of collecting unbiased WTA or WTP values from participants.

2.3.2 Designing a Contingent Valuation Experiment

Contingent valuation (CV) uses "survey questions to elicit people's preferences for public goods [or any non-market goods] by finding out how much they would be willing to pay for specific improvements them", or how much they would need to be compensated for the loss of a particular good (Mitchell & Carson, 1989, p. 2). These willingness to pay (WTP) and willingness to accept (WTA) values can be elicited using a variety of methods, though all techniques involve three key procedures (Mitchell & Carson, 1989): (1) Designing a hypothetical market model, (2) eliciting WTA or WTP data from participants, and (3) collecting characteristics data from respondents.

2.3.2.1 Creating a Hypothetical Market Model

The researcher must construct a hypothetical yet detailed and plausible market model, which is read aloud to the respondent or communicated through an easy to understand description. The market model should describe the "good which is being valued, the baseline level of provision, the structure under which the good is provided, the range of available substitutes, and the method of payment" (Mitchell & Carson, 1989, p. 3). For example, if a researcher wanted to evaluate the benefit to society of improving the water quality of a local lake, then they may design a contingent valuation scenario to describe this change, in order to elicit valuation estimates from a sample of local residents.

The good being valued is the change in water quality of the lake, so the researcher may want to describe the lake, its various recreational uses, how many people visit the lake per year, and the native fish and bird species which reside in the lake. The baseline level of provision in this scenario may be that the lake currently contains a given level of nitrogen and phosphorus which sometimes create algae blooms rendering the lake unusable by residents and unsuitable for fish and bird life for three months out of the year. The change in water quality, which is the non-market good under valuation, can then be described in terms of a particular reduction in nitrogen and phosphorus levels, which will result in the lake remaining usable all year round. The provisional rule (i.e. the structure under which the good is provided), could be that if the average WTP of participants for a system to remove excess nitrogen and phosphorus from the water is greater than the cost of the system, then the system will be purchased. Available substitute goods, which should also be described to participants, could include alternative methods of controlling the algae blooms. The method of payment for the good may be a yearly fee of $10 per household in the area surrounding the lake.

Designing a market model to seem as realistic as possible is important since participants will be making their valuation within this market context (Bateman et al., 2002). If the hypothetical scenario closely mimics the structure of a real market, then the thought process that participants go though when making a valuation statement should be similar to how they value goods in a real market

environment, resulting in stated WTP and WTA values which reflect respondent's true value for the good.

The question or questions which are used to elicit participant's WTP or WTA for a good are the second important design feature of a contingent valuation scenario. Several commonly-used elicitation mechanisms will now be described.

2.3.2.2 Choosing an Elicitation Mechanism

After "presentation of the valuation scenario, the provision mechanism and the payment mechanism, respondents are asked questions to determine how much they would value the good if confronted with an opportunity to obtain it" (Bateman et al., 2002, p. 135). This step involves obtaining stated WTP or WTA values from the respondent using questions which are clear and easy to understand but do not themselves create a bias the respondent's answers. A WTP question may ask a respondent 'would you be willing to pay $10 to gain Good A', while a WTA question may ask them 'would you be willing to accept a payment of $10 for forgo the provision of Good A'. To these types of questions, known as *single-bounded dichotomous choice* questions, participants would answer either *yes* or *no*, providing the experimenter with an indication as to whether their true WTP or WTA lay above or below $10. An obvious disadvantage of this type of question is that it only provides one upper or lower limit to a participant's valuation of a good, and therefore cannot provide a direct measure of their maximum WTP or WTA.

While it may seem plausible to directly ask an individual how much they are willing to pay for a particular good, "respondents often find it difficult to pick a value out of the air... without some form of assistance", and this can lead to inaccurate WTP estimates (Mitchell & Carson, 1989, p. 97). Researchers have therefore developed a number of different elicitation-mechanisms which aim to extract an individual's true maximum WTP or minimum WTA.

Bateman et at. (2002) outlines five of the most commonly used CV elicitation mechanisms: *open ended questions*, *the bidding game*, *payment cards*, the *single-bounded dichotomous choice* method, and the *double-bounded dichotomous choice* method. Each of these techniques will now be discussed.

Open Ended Questions

The open ended direct elicitation format asks a respondent a question such as: 'What is the maximum amount per year that you would be prepared to pay to improve the water quality of the Waikato River, in the way previously described?' The benefit of the *open ended* question style is that it does not provide the respondent with any indication as to what the value of the change might be, so the presence of an *anchoring bias* is minimised (Bateman et al., 2002). However, because people often find it difficult to pick a value out of the air with no tangible basis for the valuation, *open ended* questions tend to result in large non-response rates or protest bids (Mitchell & Carson, 1989).

The Bidding Game

The *bidding game* was one of the most widely used elicitation formats in the 1970s and 1980s (Bateman et al., 2002, p. 138), and asks respondents a question such as: 'Would you be willing to pay $5 per year for an improvement in the water quality of the Waikato River, as described previously?' If the response is 'Yes' then the bid is increased by a given amount until the response becomes 'No', and this is considered the individual's maximum WTP for the change (Bateman et al., 2002).

The simplicity of the *bidding game*, and the iterated nature with which questions are asked is thought to "capture the highest price consumers are willing to pay, thereby measuring the full consumer surplus" resulting from the proposed change (Mitchell & Carson, 1989, p. 99). A major disadvantage with the *bidding game* however, is that the value of the starting bid, and the rate at which the bids increase, may be interpreted by respondents as suggesting a possible value for the good, creating an *anchoring point* bias (Bateman et al., 2002). This bias can lead to many unrealistically high WTP bids as respondents wish to avoid the socially awkward situation of saying 'No' to a change which would be beneficial to their peers, especially when individual bids are disclosed to the others within an experiment group.

The Payment Card

The *payment card* approach presents respondents with a full list of possible payment values (e.g. $1 - $100 in $5 iterations), and could ask: 'Which of the amounts listed below would best describe you maximum willingness to pay, per year, for the increase in the water quality of the Waikato River as outlined previously?' This elicitation technique was developed in an attempt to "maintain the properties of the direct question approach while increasing response rates for WTP questions by providing respondents with a visual aid" (Mitchell & Carson, 1989, p. 100). This procedure reduces the chance of an *anchoring point* bias yet still provides a context which respondents can use to help guide their valuation. However, the range of prices listed on a *payment card*, and the increment at which the price increases, can still create a valuation bias in that respondents may pick values somewhere in the middle of the range even if these values do not represent their true WTP (Bateman et al., 2002).

Single-Bounded Dichotomous Choice

The *single-bounded dichotomous choice* or referendum method presents individuals with a single Yes/No question such as: 'Would you pay $5 per year to improve the water quality of the Waikato River in the way described previously?', with the payment amount varied across the sample. This elicitation format simplifies the cognitive process for the respondent by presenting the valuation question in way which is similar to how they would decide to purchase a particular good from a supermarket (e.g. if the price is less than their true WTP then they will purchase the good). This method reduces outliers and non-responses (Bateman et al., 2002), and with no incentive to exaggerate their WTP (i.e. in order to promote acceptance or opposition of a particular policy decision), this elicitation method is considered *incentive compatible. Incentive compatibility* means that it is in the respondent's strategic interest to answer 'Yes' if his WTP is equal to or greater than the price level, and answer 'No' if it is not (Hoehn & Randall, 1987).

The drawback of *single-bounded dichotomous choice* elicitation technique is that it does not directly specify the maximum WTP of an individual. Rather it only provides direction as to whether their WTP is greater or less than a particular

value. Therefore, more observations are required to obtain WTP estimates that have the same level of statistical significance as those elicited using other methods (Mitchell & Carson, 1989, p. 101).

Double-Bounded Dichotomous Choice

The *double-bounded dichotomous choice* method follows the format of the previous technique by asking the respondent a Yes/No question, such as: 'Would you pay $5 per year to improve the water quality of the Waikato River in the way described previously?', with the payment amount varied across the sample. However, once the first question is answered, a follow-up question of the same format offers a second price level which is higher than the first if the previous response was 'Yes', and lower if it was 'No'.

Using this *double-bounded* elicitation technique can offer greater estimation efficiency than the *single-bound* alternative since a respondent's WTP may be captured within the lower and upper bounds of the questions (e.g. between $5 and $15). However, using the *double-bounded* question format may remove the incentive compatibility of the method since "the second question may not be viewed by respondents as exogenous to the choice situation", and *anchoring bias* and *yea-saying* may also be present (Bateman et al., 2002, p. 141).

Each elicitation format has its advantages and disadvantages, and the researcher must assess these factors in relation to the needs of their experiment and any time or monetary constraints they face. Once the elicitation mechanism is decided upon, the final stage of experimental design is to construct questions to collect characteristics data from participants.

2.3.2.3 Collecting Characteristic Data

The final step in the designing a contingent valuation experiment, is to create several follow-up or debriefing questions. According to Bateman et al. two important types of debriefing questions should be included in a contingent valuation experiment: "questions to explain *why respondents were or were not*

willing to pay for the change presented..., [and] questions to explain respondent's *views of the scenario presented*" (2002, p. 145).

Collecting data on why participants are or are not willing to pay for a change, such as the increase in water lake quality, allows the researcher to assess to the validity of individual value responses (Bateman et al., 2002). Including questions which assess a respondent's attitudes, opinions, and knowledge regarding the good being valued, can be used to verify whether a WTA or WTP statement is based on the actual preferences of the respondents, or is simply an unfounded statement. For example, if a participant states a high WTP for the proposed increase in water quality of the lake from our previous example, but the follow up questions suggests that the respondent cares little for the lake's recreational, aesthetic, and habitation value, then their stated WTP may not be an accurate estimate of their true WTP.

Collecting data on how participants viewed the CV scenario, after completing the elicitation procedure, allows the researcher to assess the credibility of the valuation results as a whole (Bateman et al., 2002). If participants found the scenario to be vague, confusing, or unrealistic, then the WTA and WTP estimates obtained through the CV experiment could be considered invalid as they may not have captured the respondent's true valuation for the good.

The aim of the CV experiment is to: (1) place participants within a plausible hypothetical scenario which mimics a real market environment, (2) elicit WTP or WTA data from participants through a mechanism which does not in itself lead to a bias in their valuation, and (3) collect characteristics data from participants to assess whether their value estimates are based on their actual preferences. If designed correctly, the contingent valuation method can be used to estimate the true value that the population, from which the sample was drawn, places on the non-market good.

2.3.3 Evaluating the Validity and Reliability of CV Data

The contingent valuation method is controversial "because it involves asking individuals directly about monetary valuation related to given hypothetical

changes in the provision of an amenity", and some economists believe such a direct approach cannot produce meaningful value estimates (Veisten, 2007, p. 205). Venkatachalam notes that the "major criticism of results of CVM revolves mainly around two aspects, namely, (a) validity and (b) reliability" (2004, p. 90), where "validity refers to the 'accuracy' and reliability refers to 'consistency' or 'reproducibility' of the CV results (Kealy, Montgomery, & Dovidio, 1990).

Opponents to the CVM argue that the validity and reliability of such experiments are questionable and that WTP estimates are potentially random responses which do not reflect any true underlying preference of the participant (Veisten, 2007). To dispel these speculations, proponents of the CVM have developed various ways to test the validity and reliability of an experiment, in order to justify the value estimates they produce.

2.3.4 Validity

The validity of a CV study is essentially the degree to which this particular methodological approach is able to measure the true economic value that individuals place on a good. According to the literature there are two measures of validity that can be applied to the CVM: convergent validity and theoretical validity (Mitchell & Carson, 1989).

Convergent validity "concerns the correspondence between a measure and other measures of the same theoretical construct" (Mitchell & Carson, 1989, p. 204). Essentially, convergent validity requires that the estimated monetary value of a non-market good should be consistent regardless of whether the value was estimated through the CVM or some other valuation method. For example, comparing the value of organic cotton estimated through the CVM to the value for organic cotton calculated through hedonic price estimation could provide a measure of the convergent validity of these two measures. If the values estimated from the two different methods converge then this would imply they are both valid in terms of convergent validity.

Theoretical validity, on the other hand, involves "assessing the degree to which the findings of the study are consistent with theoretical expectations" (Mitchell & Carson, 1989, p. 206). Testing the theoretical validity of a CV experiment often involves regressing some measure of an individual's WTP

against a group of independent variables which should, according to economic theory, explain at least a portion of the respondent's WTP. The "size and sign of the estimator coefficients are then examined and judged to be consistent or inconsistent with theory" (Mitchell & Carson, 1989, p. 206).

2.3.5 Reliability

Reliability is concerned with the ability of the particular contingent valuation experiment to produce valuation estimates which reflect the true preferences of participants by minimising sources of random variation in valuation results. Put simply, "reliability refers to the extent to which the variance of the WTP amounts given by respondents in a contingent valuation survey is due to random sources, or 'noise'" (Mitchell & Carson, 1989, p. 211).

Variation in respondent's stated WTP can arise from three principle factors: (1) real underlying differences in participant's valuation of the good, (2) variation which arises due to the design of the contingent valuation experiment, and (3) variation caused by the sampling procedure used to select participants (Mitchell & Carson, 1989, p. 211). The greater the reliability of a contingent valuation experiment, the less the variation in stated WTP is caused by random sources such as experiment design or the participant sample. Therefore, reliable WTP estimates vary only as much as the true differences in the underlying preference of participants.

The reliability of an experiment can be assessed by a test-retest method, where WTP estimates are elicited from a group of participants through an identical CV procedure at two different points in time. A strong correlation between the value estimates for each participant at the two different points in time would indicate the CVM is reliable since it would appear that variation in WTP is caused primarily by differences in the true preference of participants. A very low correlation between participant's WTP over the two experiments would indicate that variation in the value estimates is caused largely by factors relating to the experimental procedure.

Reliability requires that, in repeated measurements, (a) if the true value of the phenomenon has not changed a reliable method

> *should result in the same measurement (given the method's accuracy) and (b) if the true value has changed a reliable method's measurement of it should change accordingly.*
>
> (Loomis, 1990, p. 79)

In a national transportation risk survey, Jones-Lee, Hammerton, and Philips (1985) collecting valuation data using CV, then re-interviewed a subgroup of their original sample a month later and found no significant difference in the response of the individuals. Loehman and De (1982) conducted a test-retest CV experiment to collect valuation data from 45 college students on their valuation of an air quality control policy. The correlation between participant's original WTP amounts and their stated WTP three weeks later was very high ($r = 0.86$), indicating the reliability of their experiment was strong.

Using the test-retest method to measure an experiment's reliability is very costly however, especially when the survey involves the general population, as locating the same group of respondents and convincing them all to participant may be difficult (Mitchell & Carson, 1989, p. 212). Therefore, to test the reliability of a study, most researchers aim to show that the valuation responses they collect are not purely random. This can be done "by obtaining a respectable R^2 when regressing WTP on a set of independent variables, since the higher the R^2, the lower the random portion of the WTP response variance" (Mitchell & Carson, 1989, p. 213). Testing for reliability using regression analysis differs from testing for *theoretical validity* in that it does not require that the independent variables are theoretically connected to the dependent variable. Mitchell and Carson (1989) suggests an R^2 lower than 0.15 including only a few key variables, would throw the reliability of experiment into question.

2.4 Willingness to Pay and Willingness to Accept Measures of Value

Contingent valuation studies "usually consider two different questions in order to assess a respondent's valuation of a given good or service: (i) the maximum buying price or willingness-to-pay (WTP) and (ii) the minimum selling price or willingness-to-accept (WTA)" (Schmidt & Traub, 2009, p. 229). These WTP and WTA values are collected from a representative sample of a population and then analysed in order to make inferences about how that population as a whole values the non-market good under analysis (Bateman et al., 2002). Many economists suggest that these two measures of an individual's value should be similar for most goods, with any small differences being attributed to income effects (Knetsch & Sinden, 1984; Randall & Stoll, 1980; Willig, 1976). However, "repeated experimentation has shown that values of WTP and WTA for the same good can be vastly different", and that this disparity cannot be attributed entirely to an income effect (W. L. Adamowicz, Bhardwaj, & Macnab, 1993, p. 416). The ratio of "WTA:WTP lies between 3:2 and 3:1 for private goods like mugs, chocolate bars or hockey tickets but takes on much higher values for publicly provided goods" (Bischoff, 2008, p. 283). Understanding the cause of this disparity is important to ensure that contingent valuation studies are able to produce realistic and relevant estimates of increases or decreases in society's welfare associated with proposed changed in a non-market good.

This section of literature review aims to define the concepts of WTP and WTA, and explain what these two measures of value are actually quantifying in terms of people's preferences. Following Bateman et al (2002), indifference curves will be used to define the concepts of WTP and WTA in terms of compensating variation, equivalent gains and equivalent losses. Lastly, the concept of a disparity between WTA and WTP measures of value will be introduced, leading onto the following section of literature review.

2.4.1 Measuring value in terms of WTP and WTA

Carson (1989) defines *willingness to pay* (WTP) as "the amount of money an agent would be willing to give up to obtain a change and still be as well off as his previous entitlement", while *willingness to accept* (WTA) "is the amount of

money which would have to be given to an agent, with a specified entitlement, to forgo a change and still be as well off as if the change had occurred" (p.25). According to this definition, WTP would be a measure of the maximum price an individual would pay to receive a new t-shirt, and WTA would be the level of payment offered to an individual, at which they would be indifferent between receiving a free t-shirt or the monetary compensation. Some studies however have simply defined WTP as an individual's maximum 'buying' price, and WTA as an individual's minimum selling price (Kahneman, Knetsch, & Thaler, 1991). Defining WTP and WTA values as being *buying* and *selling* prices infers that some participants have ownership of the good (i.e. the sellers) and others do not (i.e. the buyers), while Carson's (1989) perhaps more generalised explanation does not specify any kind of initial ownership. The exact specification of what WTP and WTA are measuring is therefore largely dependent on the experimental context put forward by the researcher. However, WTP is always a measure of how much an individual is *willing to pay* to accept or prevent a change from occuring, and WTA is always a measure of how much money an individual would *need to be paid* to accept forgo a change in provision. Whether or not this change in provision is the gain of a good or a loss of a good depends on the experimental situation.

Essentially WTP and WTA are both trying to measure how much monetary value an individual places on a good (e.g. a mug), or a particular attribute of a good (e.g. added branding on a mug). However, the cognitive process that a respondent goes through when answering a WTP elicitation question is quite different to the process of making a WTA judgement.

Bateman et al. explains the concepts of WTP and WTA through the use of indifference curves and changes in an individual's level of utility (2002, p.24). The authors also offer a brief explanation as to the cause of the WTA-WTP gap, drawn from the relevant literature. These WTA-WTP gap theories are addressed in further detail later in the literature review.

2.4.2 Value Measures and Indifference Curves

The two indifference curves (I and I') in Figure 1 represent the preferences of an individual between consuming various quantities of a composite for all

private goods, measured on the Y-axis, and quantities of a public good, measured on the X-axis under the constraint of a fixed level of income (Bateman et al., 2002). The indifference curves I and I' link all the possible combinations of the two goods between which the individual is indifferent. Each indifference curve "can be thought of as corresponding to a level of welfare, utility or well-being, with I' corresponding to a higher level... [since as] the indifference curves move up and to the right, the welfare of the individual increases" (Bateman et al., 2002, p. 25).

Since the Y-axis is measuring an individual's expenditure on a composite of private goods, which have a monetary value attached to them, it is possible to use these indifference curves to place an inferred value on changes in the public good, which does not have a monetary value attached, by looking at how private good expenditure changes relative to the public good. The true WTP and WTA of an individual is based on the standard economic principal that people are rational decision makers and will allocate their income between private good expenditure and public good expenditure in a way that will maximise their personal utility function (Randall & Stoll, 1980). While Bateman et al.'s example uses public good expenditure on the x-axis, the theory is the same for any non-market good, provided private good expenditure is a non-perfect substitute for the good being valued on the x-axis.

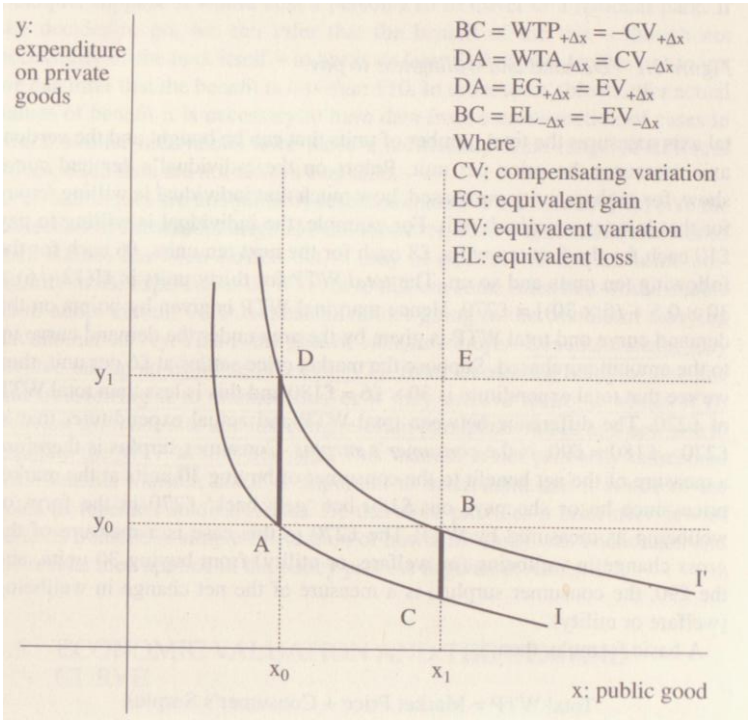


Figure 1. Measure of Change in Human Welfare

(Bateman et al., 2002, p. 24)

From Figure 1 we can identify "four measures of the value of a change in the quantity of a public good" known as compensating variation for the increase in a good, compensating variation for the decrease in a good, equivalent gains, and equivalent losses (Bateman et al., 2002, p. 25). These measures are all dependent on the initial private and public good consumption level, the shift in consumption along a single indifference curve, and also the expansion or contraction of the indifference curve itself.

Compensating Variation for a Gain

Compensating variation for an increase in a public good is measured by the decrease in private good expenditure that maintains the individual's original level of utility (Bateman et al., 2002). Suppose that an individual's consumption of the two goods begins at position A, with y_0 consumption of the private good and x_0 consumption of the public good. If consumption of the public good is increased from x_0 to x_1 so that the individual is now at position C on their

indifference curve, the decrease in private good expenditure (BC) is the amount of variation that perfectly compensates for the increase in the public good.

To illustrate the concept of compensating variation for a gain, suppose an individual with a fixed level of income was asked how much they would be willing to pay to improve the water quality of a local lake so that it became safe enough to swim in. It is assumed that the individual would receive a benefit from the additional recreation value associated with a 'clean lake', so to maintain the same level of overall utility the individual would be willing to reduce expenditure on private goods. The individual's true WTP would represent the exact reduction in private good expenditure that would maintain the original level of utility, given the improvement in the water quality of the lake.

Compensating Variation for a Loss

Compensating variation for a reduction in a public good is measured by the increase in private good expenditure which maintains the individual's initial level of utility given a decrease in the public good (Bateman et al., 2002). Assume that an individual begins at position B on indifference curve I', consuming x_1 level of the public and y_0 level of private good. If consumption of the public good was to decrease from x_1 to x_0 then, to maintain the same level of utility, the individual would increase expenditure on private goods from y_0 to y_1, shifting to position D on the indifference curve. Therefore, the compensating variation for the loss of the public good is equal to DA.

To illustrate compensating variation for a loss of a public good, suppose that an individual already has access to a lake with 'clean water' and was asked how much money they would need to be given to accept degradation of the lake to a point where it was considered unsafe for recreational use (WTA). The individual would experience a reduction of utility from the loss of the public good, but gain utility from the money they are being offered. The true WTA value of the individual would be the exact amount of money they would need to be paid to ensure their final level of utility was the same as if the lake remained clean.

Equivalent Gain

Equivalent gain is a measure of an increase in private good expenditure that would extend an individual's level of utility by the same amount as a given increase in public good consumption (Bateman et al., 2002). Assume that an individual begins at position A on indifference curve I, consuming y_0 of the private good and x_0 of the public good. If the individual's consumption of the public good was to increase to x_1 while consumption of the private good remained constant, then their indifference curve would shift from I to I' and they would now be stationed at position B. Using changes in private good consumption to create an equivalent gain in utility would require private good expenditure to increase from y_0 to y_1, holding public good consumption constant. Therefore the equivalent gain from an increase in the public good is measured by DA in Figure 1, and is the same as compensation variation for a loss, or WTA.

Again, using the hypothetical lake as an illustration, the equivalent gain measure of value for an increase in water quality would be the level of payment that would increase the individuals' utility by the same amount as the proposed increase in the public good.

Equivalent Loss

Equivalent loss is similar to equivalent gains, but is used to calculate the decrease in private good expenditure that would reduce an individual's utility by the same amount as the loss of a public good (Bateman et al., 2002). If an individual begins at position B on indifference curve I', consuming x_1 level of the public good and y_0 level of private good, then a decrease in the public good to x_0, holding private good expenditure constant, would reduce their level of utility, shifting their indifference curve to position I. To generate an equivalent loss of utility while holding public good consumption constant, expenditure on private goods would need to be reduced by the amount BC.

For example, if an individual had access to a lake with clean water which was then polluted to a point where it was no longer fit for recreational use, this loss of a public good would result in a reduction of that individual's utility,

holding all else constant. The equivalent loss measure of value for the degradation of the lake is the decrease in private good expenditure that would reduce the individual's utility by the same amount as the proposed reduction in the public good. As shown on the Figure 1, this measure is equal to the WTP of the individual.

Table 1. Summary of Valuation Measures

Gain of a Public Good	**Loss of a Public Good**
Compensating Variation (WTP) = BC	Compensating Variation (WTA) = DA
Equivalent Gain (WTA) = DA	Equivalent loss (WTP) = BC

To summarise Bateman et al.'s (2002) WTA-WTP comparison: the measure of 'value' for a non-market good that will be elicited from an individual will depend on whether the elicitation question is phrased in terms of a *gain* or a *loss*, and whether the respondent is being asked to accept or forgo the change. The four measures of value which can be elicited from participants through a contingent valuation experiment, and an example of a question to elicit each type of measure are as follows:

1) Willingness to pay for a gain: Compensating variation: BC
   E.g. 'How much would you be willing to pay for an improvement in the water quality of your local lake?'

2) Willingness to pay to prevent a loss: Equivalent loss: BC
   E.g. 'How much would you be willing to pay to prevent a decrease in the water quality of your local lake?'

3) Willingness to accept to forgo a gain: Equivalent gain: DA
   E.g. 'How much would you need to be paid to forgo an improvement in the water quality of your local lake?'

4) Willingness to accept to accept a loss: Compensating variation: DA

E.g. ‘How much would you need to be paid to accept a decrease in the water quality of your local lake?’

The critical feature to note from the Bateman et al.’s explanation is that that no matter how the question is posed (e.g. in terms of a gain or a loss), WTP always quantifies the amount BC, WTA always quantifies the amount DA, and since DA is greater than BC, WTA is expected to be slightly greater than WTP.

While willingness to pay for a gain and willingness to accept to accept a loss are both trying to estimate how much value an individual places on a non-market good or an attribute of a good, these two measures will produce slightly different results. Economic theory, as expressed by Bateman et al., explains this difference in terms of an *income effect* where participants asked for their WTA experience a higher effective income that those asked for their WTP, and this difference in point of reference results in the WTA-WTP gap. However, this income effect is expected to be insignificant in most contingent valuation scenarios, so the WTA-WTP gap, according to economic theory, should also be insignificant (Bateman et al., 2002, p. 25).

In the following section the WTA-WTP disparity will be examined in greater detail, addressing the standard economic explanation for the gap (the income effect), and explaining why this traditional theory fails to explain the often substantial differences between value measures.

2.5 **The WTA-WTP Disparity**

According to standard economic theory, the amount of private good expenditure that an individual would be willing to give up to gain a public good (Compensating Variation for a Gain: WTP) is expected to be slightly less than the amount the individual would need to be paid in order to give up the same public good (Compensating Variation for a Loss: WTA), as shown by Bateman et al. (2002) in Figure 1. This disparity is due to differences in the respondent's initial level of welfare, and is known as the *income effect* (Randall & Stoll, 1980; Willig, 1976).

According to the *income effect*, respondents asked for their WTA have a higher level of effective income than those asked for their WTP, since WTA participants hold property rights for the good they are valuing and this raises their initial level of welfare (Randall & Stoll, 1980). This difference in initial point of reference creates the WTA-WTP gap because 'wealthier' individuals would typically be willing to pay more for an increase in the public good, and would demand more compensation for a decrease in the public good (Bateman et al., 2002; Mankiw, 2007). The income effect can be clarified using indifference curves of an adapted diagram from Bateman et al. (2002, p. 25): Figure 2.

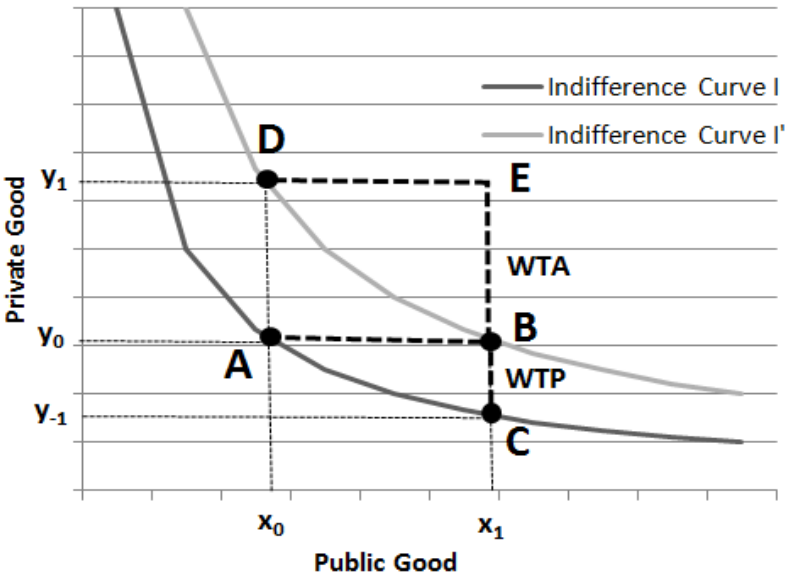


Figure 2. The Income Effect

(Adapted from Bateman et al., 2002, p. 24)

Following the reasoning outlined in section 2.4.2, the compensating variation for a gain of the public good from x_0 to x_1, or WTP, is equal to BC since the individual moves from point A to point C along indifference curve I. The compensating variation for a loss of the public good from x_1 to x_0, or WTA, is equal to EB since the individual moves from point B to point D along indifference curve I'. An individual's WTA (EB) is clearly larger than an individual's WTP (BC), following the logic outline by Bateman et al. (2002).

However, if we were to measure the compensating variation for the gain of a public good for an individual who began at the higher welfare level of indifference curve I', then their WTP would differ from that of an individual starting at the lower level of welfare of indifference curve I. For example, assume that an individual begins by consuming y_1 of the private good rather than y_0, and x_0 of the public good, meaning that they are at point D on the higher indifference curve I'. Their compensating variation for an increase in the public good from x_0 to x_1 would be EB since the individual would move from point D to point B along their indifference curve. Since EB is greater than BC, it is clear that an individual starting at a higher initial level welfare is willing to pay more for the public good simply because they are wealthier, or as Bateman et al. state: "If an individual is richer, he can afford to spend more in order to increase the public good" (2002, p. 26).

Note also than the WTP of an individual starting at the higher welfare level of indifference curve I' (EB) is the same as the WTA of an individual stationed along the same indifference curve (also EB). Therefore, if measuring WTA and WTP while holding the individual's initial level of welfare constant, the two measures would converge. If the starting level of welfare was held at indifference curve I, then WTA and WTP would both equal BC, and if the starting welfare was held at indifference curve I', WTA and WTP would both equal EB. Accordingly, the difference between WTP and WTA is due to differences in the initial wealth of the respondents since some are endowed with the public good (WTA group) and others are not (WTP group).

2.5.1 Substantial WTA-WTP Disparities

The difference in WTA and WTP value measures, resulting from the income effect, should be relatively insignificant according to most economic theorists (Knetsch, 1990). Randall and Stoll (1980), as well as Willig (1976), construct theoretical 'bounds' for the size of expected WTA-WTP gaps resulting from the income effect. Willig's bounds suggested that if the proportion of the change in effect income multiplied by half of the income elasticity is less that 5%, and the change in welfare is smaller than 90% of income, then the difference between WTA and WTP measures should be 'small' (Willig, 1976).

Standard "economic theory assumes that the amount an individual is willing-to-pay (WTP) to obtain a good is approximately equal to the amount she is willing-to-accept (WTA) to relinquish the same good", with the exception of a small disparity attributed to the income effect (Sayman & Öncüler, 2005, p. 290). However, "this near equivalence prediction has been contrary to the results of most empirical work trying to elicit WTP and WTA" which has regularly identified substantial differences between the two measures of value (W. L. Adamowicz et al., 1993, p. 417).

Brown and Hammack (1973), in one of the early articles to present both WTP and WTA data, found that waterfowl hunters were willing to pay an average of $247 to be able to continue hunting, but required $1044 in compensation to sell their hunting rights. Knetsch and Sinden (1984), in another early CV experiment, reported a WTA/WTP ratio of 4 for lottery tickets. Researchers then began to analyse this unexpectedly large valuation gap by collecting WTA and WTP data from participants for a wide range of goods, including both non-market and regular market goods.

Substantial WTA-WTP disparities, much larger than the income effect would suggest, became so widespread in the stated preference literature that several authors compiled the results of many individual studies in order to perform a meta-analysis on other potential causes of the WTA-WTP gap (Horowitz & McConnell, 2002; Sayman & Öncüler, 2005). Horowitz and McConnell's meta-analysis included the results of 45 individual studies, which comprised 201 WTA-WTP experiments (2002, p. 428). The authors found that

significant WTA/WTP ratios were observable in a majority of the experiments they examined, also noting that most of the experiments' authors had "remarked that the WTA/WTP ratio is much higher than their economic intuition would predict" (Horowitz & McConnell, 2002, p. 426). The average WTA/WTP ratio for the 201 experiments included in Horowitz and McConnell's analysis was 7.17, and ranged from 1.95 to 10.41 depending on the type of good being valued (2002, p. 433).

Sayman and Öncüler's (2005) meta-analysis included 164 WTA/WTP ratios collected from 39 individual studies. The average ratio from their sample was 7.1 and ranged from 0.14 to 113 (Sayman & Öncüler, 2005, p. 300). Ratios of less than 1 were observed 7 times in their data set, indicating that the selling price in that experiment was lower than the buying price.

Long before Horowitz and McConnell (2002) and Sayman and Öncüler (2005) conducted their respective meta-analyses, it was clear that WTA was significantly different from WTP in a majority of the stated preference literature, and that this difference could not be explained by convention economic theory (i.e. the income effect). Contingent valuation researchers therefore looked for other explanations as to why an individual's WTA for the loss of a good is greater than their WTP to gain the good. Most notable of these theories, each of which will now be discussed, include *weak experimental design*, the *substitution effect*, and the *endowment effect* (Mansfield, 1999, p. 220).

2.6 Weak Experimental Design

Some researchers have suggested that the large WTA-WTP gaps observed in the literature are simply the result of weak experimental design features such as using hypothetical payments, elicitation techniques that are not incentive-compatible, and a general misperception by participants about how the valuation process actually works (Horowitz & McConnell, 2002). According to this argument, experiments designed to be more realistic, "such as those with real money or incentive-compatible elicitation, will yield lower and more reasonable ratios" (Horowitz & McConnell, 2002, p. 426). Furthermore, ensuring subject anonymity and clearly explaining the elicitation procedure should also reduce the WTA-WTP disparity (Plott & Zeiler, 2005).

This following section literature review will look at the evidence regarding the *weak experimental design* explanation for the WTA-WTP gap. It will do so by reviewing various experimental investigations which dealt with three of the theory's fundamental components: (1) incentive compatibility of the elicitation mechanism, (2) ensuring respondent anonymity, and (3) providing a thorough description to participants about how the elicitation mechanism works.

2.6.1 Incentive Compatible Elicitation Mechanism

The "incentive properties of stated-preference surveys is a central debate in the valuation of public goods" (Taylor, Morrison, & Boyle, 2010, p. 198). An elicitation method is considered to be *incentive compatible* if the technique provides "respondents with incentives to reveal their true valuation" of the good under analysis (Bateman et al., 2002, p. 128). If the elicitation mechanism is not incentive compatible then the WTA and WTP data collected from respondents may be subject to various degrees of bias, and will provide an inaccurate estimate of the true value they place on the good.

Incentive compatibility is based on a range of features of the adopted methodology. Two important features are: (1) whether or not the experiment involves real monetary transactions or is purely hypothetical, and (2) the conditions under which participants are either given or denied provision of the good, known as the *provisional rule* (Bateman et al., 2002).

2.6.1.1 Hypothetical Bias

Some researchers claim that if an experiment involves purely hypothetical decision making tasks, where the values stated by participants are in no way connected to actual required payments or the provision of the good, then there is little incentive for respondents to reveal their true WTP or WTA (Bateman et al., 2002). Participants in hypothetical experiments may exaggerate their stated WTP in order to appear more generous to their peers, or to themselves (Bryan & Jowett, 2009). Another "possibility is that the answers people give to hypothetical questions are likely to be off-the-cuff or careless responses which do not reflect their true taste preferences" (Carson, 1989, p. 172).

Introducing real monetary transactions into the experiment is often considered to make the contingent valuation scenario seem more realistic to participants, and therefore increase the amount of thought that respondents put into making their valuation statements, resulting in a more accurate representation of the true value they place on the good (Carson, 1989). The difference in valuation estimates collected from hypothetical experiments versus binding choice experiments - where value statements have a binding monetary consequence – is known as a *hypothetical bias*. While there is experimental evidence to suggest that a significant difference exists between valuation data collected through a binding-choice scenario or a hypothetical scenario (Cummings, Harrison, & Rutstrom, 1995; List & Gallet, 2001; Taylor et al., 2010), there is also research to contradict this theory (Camacho, García, Georgantzís, & Sabater, 2004; Horowitz & McConnell, 2002).

A laboratory experiment conducted by Cummings et al. which tested for *hypothetical bias* found that the WTP "values elicited by the real DC [dichotomous-choice] question, which *is* incentive compatible, were significantly different from the values elicited by the hypothetical DC question..., [and] this result is robust to different private goods" (1995, p. 266).

Taylor, Morrison, and Boyle (2010) provide further evidence to support the notion of a hypothetical bias in contingent valuation experiments. The authors construct a choice experiment to elicit WTP values for various attributes of a simple market good (a t-shirt), and also for attributes of a public good (planting of

trees is a local park) under both a hypothetical scenario and a binding scenario. In the binding treatment, participants were informed that one of the choice options they are given would be selected at random, through the roll or a dice, and the particular purchasing choice they made on that question would enforced. For example, if a choice set asked participants to select between option A: a t-shirt with short sleeves and no logo for $10, option B: a t-shirt with short sleeves and a logo for $14, or option C: no purchase, and the participant selected option A, they would be required to pay $10 in return for the described t-shirt if that choice set was chosen to be binding.

Using a multinomial probit regression model, Taylor et al. estimate the marginal WTP of participants for attributes of the private good, which include whether the t-shirt had short or long sleeves and a logo or no logo, and features of the public good, which included the type of tree being planted, and the size of the tree. They find that "that marginal WTP estimates from the hypothetical treatments are much larger, and statistically different than corresponding estimates in the binding choice treatment" in the public good experiment (2010, p. 197). However, their results did not show a significant difference between WTP estimated from the hypothetical and binding treatments in the private good experiment.

Camacho, García, Georgantzís, and Sabater (2004) conduct an experiment in which hypothetical and real WTP data is elicited from participants for an improvement in the recyclability of an office table. The experiment involved 76 volunteers who were recruited from a Business Administration and Engineering course at the University Jaume I. The participants were provided with a brief description of two office tables which were identical except that table A was able to be dismantled and recycled at the end of its lifecycle, where as table B had to be sent to a landfill. Respondents were then asked which of the two tables they would prefer to purchase in seven choice tasks where the price of the recyclable table increased from 26,000 ECU (experimental currency units) to 38,000 ECU at increments of 2,000 units, while the price of the non-recyclable table remained constant at 28,000 ECU (Camacho et al., 2004). Participants WTP for the recyclability attribute of table A could be derived from the maximum price they

were willing to pay for the recyclable table before they would prefer to purchase the non-recyclable alternative. For example, if the participant was still willing to purchase table A at a cost of 32,000 ECU before they would rather purchase table B, then their WTP was 4,000 ECU which is the difference in price between the table B and table A.

In the hypothetical treatment group of Camacho et al.'s (2004) experiment, no connection was made between the valuations made by participants and the monetary compensation they received for taking part in the experiment. However, in the binding-choice treatment, participants were given an initial endowment of 40,000 ECU and a provisional rule was introduced which added an incentive compatibility component to the elicitation mechanism. Participants in this binding treatment were given instructions identical to those in the hypothetical treatment, but were informed that one of the seven choice tasks they made would be selected at random, and two participants from the treatment would need to use part of their 40,000 ECU endowment to purchase either table A or table B for the price stated on that choice line. The experimental units that participants were endowed with could be traded in at the end of the experiment for real money, so participants forced to purchase one of the tables would essentially be spending real money on the good.

Camacho et al. (2004) compare the results of their two elicitation sessions (hypothetical and binding), to determine if the incentive-compatible binding treatment produced WTP data that was significantly different from that collected from the hypothetical treatment. The authors found that, "contrary to most of the results obtained in similar studies, at a population level, there are no significant median differences between actual and hypothetical stated values of WTP" (Camacho et al., 2004, p. 313). The price range at which the average respondent was indifferent between purchasing the recyclable table or the non-recyclable table was between 34,000 and 36,000 ECUs for the hypothetical treatment, and between 32,000 and 34,000 ECUs for the binding treatment, a difference that was not found to be statistically significant (Camacho et al., 2004, p. 322).

Evidence from Meta-Analysis

An extensive analysis of the WTA-WTP experiments conducted by Horowitz and McConnell, found results similar to Camacho et al. (2004), concluding that "real experiments do not yield ratios that are significantly different from those of hypothetical experiments..., [and] thus, any claim about the suitability of hypothetical surveys must rest on evidence other than the size of the WTAWTP ratio" (2002, p. 437). However, List and Gallet's (2001) meta-analysis of 29 experimental studies provides compelling evidence in support of the existence of a hypothetical bias, finding that on "average subjects overstate their preferences by a factor of about 3 in hypothetical settings" (2001, p. 241).

List and Gallet's experiment also finds that "willingness to pay studies yield smaller hypothetical-to-actual ratios than willingness to accept studies", and that the hypothetical bias is greater for public goods than for non-public goods (2001, p. 251). The authors suggest that "subjects should be more apt to correctly state their true preferences when performing a familiar hypothetical task (WTP) rather than an unfamiliar one (WTA)" (List & Gallet, 2001, p. 248). Furthermore, similar reasoning can be used to explain their finding that hypothetical bias is greater for public goods than for private goods since participants are more familiar with valuing private goods and are therefore more likely to accurately estimate their WTP for such goods in a hypothetical experiment. This finding would suggest that in a hypothetical CV experiment, WTA will be subject to a greater upwards bias than WTP, therefore resulting in a larger WTA-WTP disparity than would be expected in binding choice experiments.

While the hypothesis that experiments involving hypothetical payments will produce significantly different results that those involving real payments seems plausible, the experimental evidence on this theory is mixed. It does appear however, that any hypothetical bias within an experiment tends to be smaller for valuation tasks which respondents are familiar with, such as stating their WTP for ordinary private goods.

2.6.1.2 The Provisional Rule

Assuming an experiment involves real monetary transactions, a second important feature of the elicitation mechanism is the provisional rule which

determines whether or not a participant receives the good they are valuing and determines the price that they are required to pay for it. An example of a provisional rule could be that if a respondent's WTP is greater than a predetermined or randomly generated value, then the participant must purchase the good for their stated value. It is important to insure that the provisional rule does not encourage participants to employ strategic response behaviour in their valuation (e.g. by stating values higher or lower than their true WTA or WTP in an attempt to effect the provision of the good).

Mitchell and Carson (1989) suggest that if respondents believe that provision of the good is a guaranteed outcome and that they will have to pay the amount they state, then there is a clear incentive to understate their true value for the good. Conversely, if respondents believe that their stated value will influence the provision of good (e.g. in a referendum vote), but the amount of money they will be required to pay is unrelated to their bid, then there is a clear incentive to overstate their true valuation of the good (Bateman et al., 2002, p. 128). Whether or not the elicitation mechanism encourages understatements or overstatements of respondents' true WTA or WTP is dependent on the particular provisional rule employed, but both of these biases are undesirable in contingent valuation studies.

Taylor states that a provisional rule "is incentive compatible if there are no incentives [for participants] to misrepresent [their] true preferences", meaning there is no strategic reason for respondents to produce a biased value statements (1998, p. 133). It is suggest that true value estimates can only be elicited if respondents believe that provision of the good is contingent on their stated value, and that they will be required to pay the amount they state (Mitchell & Carson, 1989).

2.6.2 Subject Anonymity

Failing to ensure subject anonymity is argued by some researchers to affect the behaviour of participants involved in an experiment and impact on their stated WTA and WTP for a particular good (Fremling & Posner, 1999; Hoffman, McCabe, Shachat, & Smith, 1994; Plott & Zeiler, 2005). The basis of this argument is that participants making decisions without anonymity are more likely

to think about how other people will view their decision and may therefore provide what they think is a 'socially acceptable' response, rather than a response based on their true preference.

Hoffman et al. show that the degree of anonymity given to a participant effects the outcome of a Dictator Game in which some participants are provided with an initial endowment (the dictator) and must decide how much to share with a fellow participant who did not receive any endowment (1994). Theory would suggest that the most economically efficient decision for the dictator would be to not share any of their endowment, as there is no adverse repercussion for doing so. However, it is a common feature in Dictator Game experiments for participants to make economically inefficient decisions and share a proportion of their endowment (Cherry, Frykblom, & Shogren, 2002; Oxoby & Spraggon, 2008) Hoffman et al. run a number of experiments, providing dictators with varying degrees of anonymity (1994). They conclude that "the more anonymous the game, the smaller the "dictator's" gift", suggesting that participants are less self-conscious about their decisions when full anonymity is ensured, and provide a more accurate representation of their true preference in line with standard economic theory (Hoffman et al., 1994, p. 29).

Fremling and Posner (1999) hypothesis that if valuation decisions are not made anonymously then participants may be concerned with how other people view their decision, and may therefore make 'socially acceptable' decisions rather than economically efficient ones. This hypothesis would explain the results of Hoffman et al.'s (1994) Dictator Game experiment, and could also offer a partial explanation for the WTA-WTP disparity in buying and selling experiments.

> *For example, talented and successful bargainers tend to sell high and buy low. Therefore, if a subject wishes to be known by other subjects or the experimenter as a talented bargainer, he might adjust his behaviour accordingly even if the elicitation device does not reward that type of behaviour*
>
> (Plott & Zeiler, 2005, p. 538)

If what Plott and Zeiler (2005) suggest is an accurate depiction of a respondent's thought process during a WTA/WTP experiment, then ensuring

participant anonymity would eliminate this kind of strategic behaviour, resulting in a stated value which is closer to participant's true values value for the good, whereby reducing the WTA-WTP gap.

2.6.3 Participant Learning

The market environment in which CV experiments place participants would be considered unfamiliar by most people who would have limited or no experience making valuations under such conditions. As List & Gallet (2001) hypothesised, people produce more accurately statements of their true preference when asked familiar type of valuation questions (WTP rather than WTA) for goods which they are also more familiar with (private goods rather than public goods). Accordingly, the average participant would find it difficult to reveal their true preference during a CV experiment, simply because they are unfamiliar with the market situation and the elicitation mechanism being used. Also, since participants may find it easier to answer a WTP question than a WTA question, the former measure of value may be closer to their true valuation of the good while the latter is subject to varying types of bias. This is thought by some researchers to be the true cause of the WTA-WTP disparity.

Researchers often use repeated trials in contingent valuation experiments in an attempt to reduce the disparity between WTA and WTP measures. The rational for this being "that by allowing respondents to `practice' in the (often unusual) market situations in which experiments place them, they are given the opportunity to refine their responses to more accurately reflect their preference" (Morrison, 2000, p. 57). Allowing for participant learning is expected to increase how accurately WTP and WTA values capture a respondent's true preference towards a good, and since these two measures should theoretically be identical (Willig, 1976), doing so would eliminate the disparity commonly observed in the literature.

Coursey, Hovis, and Schulze (1987) conducted an experiment to analyse how participant learning impacts on WTA and WTP measures elicited using a Vickery Auction for a bitter-unpleasant taste experience (tasting sucrose octa-acetate). The WTP question asked participants how much they were willing to pay to avoid tasting the product, while the WTA question asked participants how

much compensation they required to taste the bitter product. The researchers elicited values from participants over 10 trials, and found that "WTA and WTP tend to converge in a mature market setting..., [a result] consistent with economic theory" (Coursey et al., 1987, p. 688). This result suggested that as participants became more familiar with the elicitation process, both WTA and WTP statements converged on the respondent's true valuation for the good.

Shogren, Shin, Hayes, and Kliebenstein (1994), in an experiment involving ordinary private goods which participants were familiar with (mugs and candy bars), found that WTA and WTP measures converged over repeated trials, eliminating any statistically significant difference between the two measures of value. Shogren et al.'s (1994) second elicitation experiment involved valuing a reduction in the likelihood of a sandwich being contaminated with various pathogens. Their results from this experiment also found that the very large WTA-WTP ratios observed in the first elicitation trial (34 for the 4^{th} pathogen type) was reduced substantially over multiple trials (to a ratio of 3), however, participant learning was unable to completely eliminate the disparity.

In a valuation experiment involving chocolate bars, Morrison (2000) "found that WTA and WTP estimates of value not only do not converge over repeated trials, but [there was] no evidence to suggest that the disparity even decreases" (Morrison, 2000, pp. 62-63). Morrison's experiment also controlled for the substitutability of the good, another explanation for the WTA-WTP gap proposed by Hanemann (1991), and therefore offers compelling evidence against the need for repeated elicitation trials.

The effect on the WTA-WTP gap of allowing participants to become familiar with the elicitation method remains unclear. While some experiments have found that running multiple elicitation trials can reduce or eliminate the disparity (Coursey et al., 1987; Shogren et al., 1994), others have not observed this effect (Morrison, 2000). Furthermore, in a reassessment of Coursey et al.'s (1987) results, conducted by Gregory and Furby, it was noted that the sample size of Coursey et experiment was quite small, and the convergence of WTA and WTP "depends upon inclusion of... suspicious outlying groups" (1987, p. 285).

2.6.4 Simultaneously Controlling for Weak Experimental Design Features

These three features of *weak experimental design* described previously – using a non-incentive compatible elicitation mechanism, failing to ensure subject anonymity, and not providing participants with adequate training in the elicitation procedure – are argued by various commentators to encourage respondents to state biased WTA and WTP values which do not accurately reflect their true valuation of the good, leading to the commonly-observed disparity. Plott and Zeiler (2005, p. 530) note that while past experimenters had controlled for one or two weak experimental design features, none had controlled for all of these design features at once, and consequently designed an experiment to fill this gap in the literature.

Plott and Zeiler's (2005) paper provided evidence in support of the weak experimental design explanation for the WTA-WTP gap. The authors construct an experiment which simultaneously controls for all experimental design features which they identify from the literature as being a possibly cause of the WTA-WTP. Their proposed *strong* experimental design: (1) uses an incentive compatible elicitation technique (Becker–DeGroot–Marschak method - BDM), (2) provides training to participants on how the elicitation procedure works, (3) offers several paid practice elicitation rounds, and (4) ensures total anonymity for participants. The researchers use this experimental design to elicit WTA and WTP values from 74 participants for a simple market good, a mug, which other experimenters have regularly used to identify and analyse the WTA-WTP gap (Plott & Zeiler, 2005). The authors find that no significant WTA-WTP gap is observable when taking into account these four experimental features, suggesting that a weak experimental design may be the cause of the commonly observed disparity (Plott & Zeiler, 2005, p. 540).

However, Plott and Zeiler's (2005) experiment only involved a simple market good (a mug), so while controlling for weak design features may remove the WTA-WTP gap for simple market goods, it is unclear whether using a similar experimental design would remove the gap for non-market goods which typically produce a much higher WTA-WTP ratio (Horowitz & McConnell, 2002). Furthermore, other experiments have controlled for weak experimental design

features by using a market for tokens (Kahneman, Knetsch, & Thaler, 1990) and have still observed significant WTA-WTP gaps, suggesting that design features are not the entire cause of the WTA-WTP disparity.

Furthermore, Horowitz and McConnell's (2002) meta-analysis of 201 WTA-WTP valuation studies found that experimental design features were not the true cause of the observed disparity. Their analysis concluded that: WTA-WTP "ratios in real experiments are not significantly different from hypothetical experiments..., incentive-compatible elicitation yields higher ratios, not lower [as would be expected]..., [and] there is not strong evidence that the ratio decreases through iteration" (2002, p. 427). The conclusions reached in Horowitz and McConnell's (2002) paper provides substantial evidence against the weak experiment design cause of the WTA-WTP gap.

2.7 **The Substitution Effect**

Hanemann (1991) disregarded weak experimental design as the source of the WTA-WTP gap, offered an alternative explanation, suggesting that the disparity may be caused by there being a lack of available substitutes for the commodity being valued. According to economic theory, when two goods are perfect substitutes for one another, "the marginal rate of substitution is constant, [and therefore] the indifference curves are straight lines" (Mankiw, 2007, p. 461). Hanemann hypothesised that if the public good being valued had easily accessible substitutes, then the indifference curve between private good expenditure (i.e. money) and the public good would tend towards a perfect linear relationship and the difference between WTA and WTP would diminish.

> *The large empirical divergences between WTP and WTA may be indicative not of some failure in the survey methodology but of a general perception on the part of an individual that the private-market goods available in their choice set are, collectively, a rather imperfect substitute for the public good under consideration.*
>
> (Hanemann, 1991, p. 646)

As demonstrated in Figure 3 (below), when the private good (i.e. money) is a perfect substitute for the public good, the WTP for an increase in the public good from point x_0 to x_1, represented by line BC, should be equal to the WTA for an identical decrease in the public good from x_0 to x_{-1}, shown by line AE. However, when the public good is not considered a perfect substitute, then the indifference curve between private good expenditure and the public good will become convex to the origin due to a diminishing marginal rate of substitution (Hanemann, 1991).

The diminishing marginal rate of substitution, based on the principle of diminishing marginal utility, suggests that when an individual is consuming a large quantity of good X they are willing to give up less of good Y to gain an additional unit of good X (Mankiw, 2007). However, when an individual is consuming a small quantity of good X, they are willing to give up a larger quantity of good Y to gain an additional unit of good X. This convex nature of the indifference curve, shown in Figure 4, means that an individual's WTP for an

increase in a public good from x_0 to x_1, shown by line BC, will be less than their WTA for an identical decrease in the public good from x_0 to x_{-1}, shown by line EA, as theorised by Hanemann (1991).

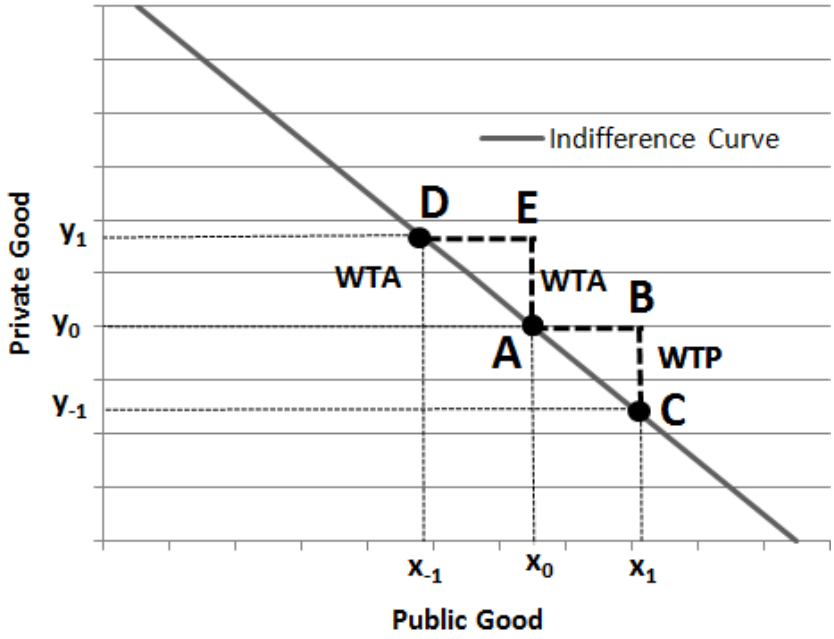


Figure 3. Indifference Curve for Public Good with a Perfect Substitute

(Adapted from Shogren, Shin, Hayes, & Kliebenstein, 1994, p.256)

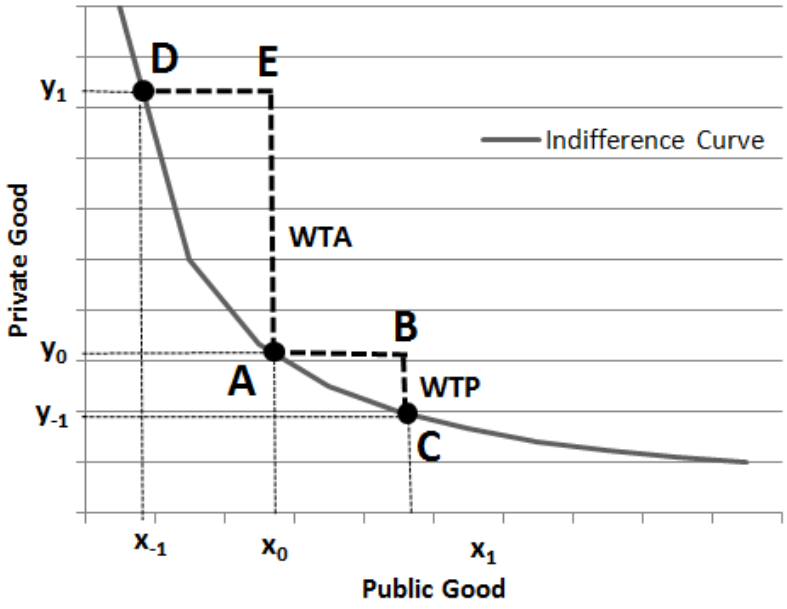


Figure 4. Indifference Curve for a Public Good with no Perfect Substitute

(Adapted from Shogren, Shin, Hayes, & Kliebenstein, 1994, p.256)

To test whether Hanemann's substitution theory could be the cause of the WTA-WTP gap, many empirical studies have looked to identify a clear relationship between the degree of substitutability of a good and the WTA-WTP gap associated with the good (W. L. Adamowicz et al., 1993; Horowitz & McConnell, 2002; Shogren et al., 1994)

In an experiment constructed by Shogren, Shin, Hayes, and Kliebenstein (1994), WTA and WTP values were elicited from 142 participants using a Vickrey second-price sealed-bid auction for an easily substitutable product, a candy bar, and a less-substitutable product, a food-borne pathogen screening procedure (1994, p. 259). In the first half of the experiment participants were asked how much they would be willing to pay (or accept) to upgrade a small piece of candy they had been given to a full-sized candy bar. The bid values were collected over a number of trials, with one trial randomly selected to be 'binding', in order to ensure incentive-compatibility. The authors used multiple trials to allow for participant learning and too see whether WTA-WTP values converged as individuals became more familiar with the elicitation procedure.

In the second half of the experiment WTA and WTP values were elicited for a pathogen-screening procedure, using the same elicitation technique that was adopted for the candy bar trials. Participants were asked how much they would be willing to pay (or accept) to upgrade a sandwich they had been endowed with, which had a standard probability of being contaminated with some kind of pathogen, to a 'stringently-screened' sandwich, which had a very low chance of being contaminated (one in a million).

Shogren et al. (1994) found that:

> *For market goods with close substitutes which are readily available in commercial outlets with minimal transaction costs (i.e., candy bars and coffee mugs),... WTP and WTA value measures converge. In contrast, for a nonmarket good with no close substitutes (i.e., reduced health risk), the value measures diverge and persist, even*

with repeated market participation and full information on the nature of the good

(Shogren et al., 1994, p. 266)

Shogren et al.'s results support Hanemann's "argument that the degree of substitutability between goods may drive the difference between WTA and WTP measures of value", though the authors did comment that further research into the robustness of their results was needed (1994, p. 266).

In their article *A Review of WTA / WTP Studies*, Horowitz and McConnell (2002) conduct a comprehensive review of WTA and WTP elicitation experiments involving 201 different products which they separate into five categories based on the type of good being valued. These categories were: Public or non-market goods, health and safety goods, ordinary private goods, lotteries, and timing goods. Summary statistics from the included experiment can be seen in Table 2 (below).

Table 2. WTA/WTP Ratio by Type of Good

(Horowitz & McConnell, 2002, p. 433)

Good	Mean RATIO	Standard error	Number of experiments
Public or non-market goods	10.41	2.53	46
Health and safety	10.06	2.28	32
Ordinary private goods	2.92	0.30	59
Lotteries	2.10	0.20	25
Timing	1.95	0.17	39
All goods	7.17	0.93	201
Unknown number of subjects	6.71	Not calculated	6

As seen in table 2 the WTA-WTP ratio is largest for *public or non-market goods* (10.41) and *health and safety goods* (10.06), while *ordinary private goods*, *lotteries*, and *timing goods* all have much lower ratios of between 1.95 and 2.92. This suggests that experiments involving goods with low substitutability (e.g. *public or non-market goods* and *health and safety goods*) observe a much greater WTA-WTP disparity than those experiments involving goods which have a higher degree of substitutability (e.g. *ordinary private goods*).

These summary statistics however do not take into account other factors which could contribute to the WTA-WTP gap. Horowitz and McConnell therefore constructed an econometric model to analyse how experimental design features, the type of good being valued, the mean WTP for the good, and the year of the experiment impacted on the WTA-WTP ratio observed in each study. Experimental design features incorporated into the regression model included whether the experiment involved hypothetical or real payoffs, which elicitation technique was used, and whether the experiment involved student or non-student subject participants. The goods being valued were classified as either *ordinary private good* or *all others*, where lotteries and timing goods were included in the *all others* category even though their WTA-WTP ratios appear closer to that of an *ordinary private good* (see table 2). The authors adopt a random-effects model which allows for covariance among multiple experiments in a single study. They calculate the model using maximum-likelihood estimation and giving more weight to experiments involving a greater number of participants.

The results of Horowitz and McConnell's regression analysis found that "non-ordinary goods have significantly higher ratios - they are typically 6 to 8 points higher than ordinary goods. This effect occurs even when we take survey design features and mean WTP into account" (Horowitz & McConnell, 2002, pp. 436-437). These results were "consistent with Hanemann's finding that the lower the substitution elasticity between a bundle of market goods and the rationed good is, the higher the WTA-WTP ratio will be" (2002, p.435).

A hypothetical experiment run by Adamowicz et al. (1993) looked at how the substitutability of a ticket to see a National Hockey League play-off game could explain variation in the WTA and WTP values elicited for the good. The experiment involved 300 first year undergraduate students from the University of Alberta. Half of the participants were informed that watching the match on television or listening to it over the radio were easily obtainable substitutes to seeing the game live, while the other half were told that no such substitutes were available. These 'substitute-available' and 'substitute-unavailable' groups were each split in two further sub-groups, with WTP values elicited from one and WTA values elicited from the other, using close-ended questions where bid amounts

varied across each sample. Adamowicz et al. found that the "difference between WTA and WTP was $8.50 (30 percent) smaller for the substitute subsample than the no-substitute subsample", suggesting that Hanemann's substitutability theory had some credibility (1993, p.425).

However, not all investigations into the substitutability effect have found clear evidence in support of the Hanemann's theory. Adamowicz et al. (1993), for example, conducted an experiment to elicit WTP and WTA values from 168 student participants using the CVM with open ended valuation questions. The item being valued was a ticket to see a particular movie at a local theatre. Movie tickets were chosen as the good to be valued because they were familiar to consumers, they represented a small portion of the participant's total income, and they had many easily accessible substitutes (W. L. Adamowicz et al., 1993, p. 419).

The aim of the Adamowicz et al.'s Movie Ticket experiment was to use characteristic data collected from each participant to judge the substitutability of the movie ticket for each individual. The degree to which the video cassette option was considered a substitute "was determined on the basis of the respondent's access to a VCR, their attitude towards VCR versus big screen theatre experiences, and their attitude towards theatre use" (W. L. Adamowicz et al., 1993, p. 419). The measure of substitutability was compared with the WTA-WTP gap of the individual under the hypothesis that the gap should be smaller for participants to whom VCR was a close substitute. Results from the movie ticket experiment were inconclusive however. Through regression analysis "the potential substitute variables were not found to be significant in explaining any observed difference between the two welfare measures [WTA and WTP]" (W. L. Adamowicz et al., 1993, p. 420). These results contradicted Adamowicz et al's (1993) Hockey Ticket experiment and offered partial evidence against the Hanemann's substitutability hypothesis.

Adamowicz et al. concluded from the results of their two experiments that "while substitutes appear to have an effect on the WTP-WTA difference, in this sample, the availability of the substitute was not sufficient to erase the significant difference between these two measures" (1993, p. 425).

2.8 The Endowment Effect

An alternate explanation for the WTA-WTP gap was put forward by Thaler (1980) who introduce the idea that an *endowment effect* may contribute to the disparity in value measures. It was suggested that "individuals asked for their WTA for a certain good will consider this good part of their endowment while individuals asked for their WTP do not..., [and] this difference in point of reference causes a disparity between WTA and WTP" (Bischoff, 2008, p. 284). While this concept appears similar to that of the *income effect*, the endowment effect theory offers an explanation why the WTA-WTP gap is often larger than standard economic theory would predict.

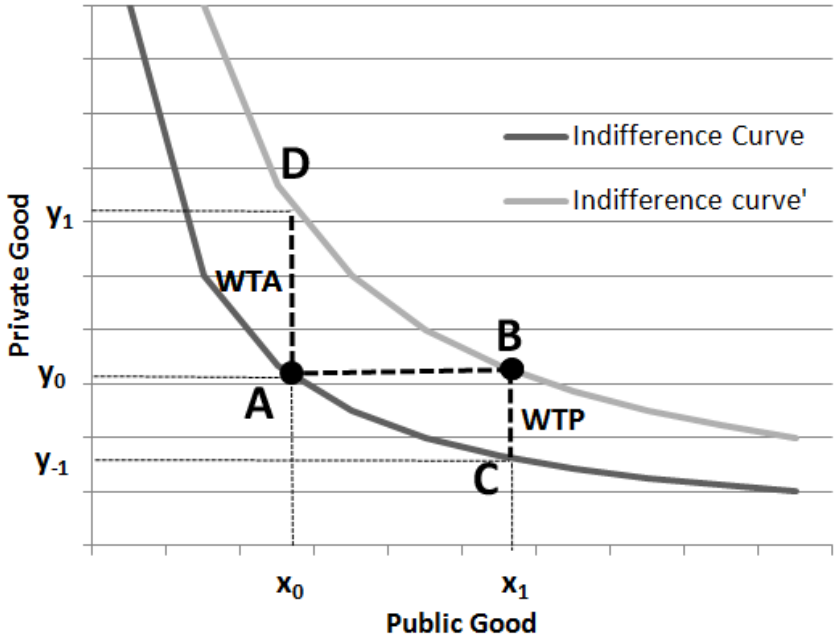


Figure 5. The Endowment Effect

Experimental investigations into the endowment effect can be broadly divided into two schools of thought: those which explain the endowment effect in terms of parting-disutility or loss aversion (Kahneman et al., 1991; Kahneman & Tversky, 1979), and those which explain the concept in terms of an unanticipated ownership utility effect (Bischoff, 2008; Loewenstein & Adler, 1995; Van Boven, Loewenstein, & Dunning, 2003).

The parting-disutility effect suggests that the "WTA–WTP disparity is caused by a disutility which the owner suffers when parting with an endowment",

resulting from the fact that the individual tends to place a greater value on a loss than a gain of equal magnitude, a concept known as loss-aversion (Bischoff & Meckl, 2008, p. 1769). The unanticipated ownership utility effect meanwhile, explains the disparity "by the fact that people get attached to goods they own but fail to anticipate the utility from feeling attached to one's endowment before being endowed" (Bischoff & Meckl, 2008, p. 1769). Therefore people place a higher value on the good once they own it and experience the additional utility that comes with ownership.

While these two explanations for the endowment effect differ on theoretical grounds, they both attempt to explain the same phenomenon: why individuals tend to place a greater value a good they possess than an identical good they do not possess. Distinguishing between a parting-disutility effect and an unanticipated ownership utility effect is difficult however, as it is possible that they both contribute to the WTA-WTP gap simultaneously.

2.8.1 The Endowment Effect and Parting Disutility

Kahneman and Tversky, in their prominent article *Prospect theory: An analysis of decision under risk*, showed that, by focusing on gains and losses associated with a choice rather than final asset provision, the "value function [of an individual] is normally concave for gains, commonly convex for losses, and is generally steeper for losses than for gains", meaning that people place a greater value on a loss than an identical gain (1979, p. 263). This finding suggests that individuals should be willing to pay more to prevent the loss of a good, than they would be willing to pay for a gain of an identical good. This theory of loss-aversion has been adopted by a group of researchers who explain the endowment effect in terms of a parting-disutility which participants experience when asked for their WTA.

Thaler (1980) suggested that "goods that are included in the individual's endowment will be more highly valued than those not held in the endowment, ceteris paribus,... because removing a good from the endowment creates a loss while adding the same good (to an endowment without it) generates a gain", and

according to *prospect theory*, the loss would be more heavily weighted in an individual's value function than the gain (p.44). Accordingly, an individual presented with a WTA elicitation question will feel as those the good being valued is part of their personal endowment, interpreting the proposed change as a loss, and will therefore place a higher value on the good than they would if they were asked a WTP question.

In Thaler's (1980) article *Toward a positive theory of consumer choice*, the author explains the thought process behind several significant marketing concepts in terms of the endowment effect and loss-aversion, since at the time there was a lack of empirical work on which to base his theory (p.45). One such example Thaler used was that of a bill which had recently been passed by United States Congress allowing shop owners to charge customers a fee to purchase products if they used a credit card. Up until that point shops had been pressured by credit card companies not to introduce such a fee, as the card companies expected such a charge to reduce the appeal of credit cards. Realising that the bill was going to be introduced, the credit card companies lobbied Congress to amend the bill so that the change was termed as a 'cash discount' rather than as a 'credit card surcharge' (1980, p.45). Thaler implied the credit card companies understood that a 'credit card surcharge' would be viewed by customers as a loss associated with using a credit card, while a 'cash discount' would be seen as a gain to using cash money. So while the change was essentially the same, the card companies believed that phrasing it in terms of a 'gain' for using cash would have less of an adverse impact on credit card use than calling it a 'loss' associated with using credit.

One of the early empirical investigations into the endowment-effect explanation for the WTA-WTP gap was Kahneman, Knetsch, and Thaler's *Experimental Tests of the Endowment Effect and the Coase Theorem* (1990). In this paper, the authors use an incentive-compatible CV experiment to determine participant's WTA and WTP for a range of ordinary market goods. These values were then compared to WTA and WTP estimates for an 'induced value' product (tokens) elicited using an identical technique, in order to isolate the endowment effect from other experiment-based effects.

Kahneman et al.'s experiment is based on the assumption that "there are some cases in which no endowment effect would be expected, such as when goods are purchased for resale rather than utilization" (1990, p.1328). Tokens are a clear example of a good which is held exclusively for resale, and Kahneman et al. hypothesise that "no endowment effect would be expected for such tokens..., thus both buyers and sellers should value tokens at the induced value they [the tokens] have been assigned", and no WTA-WTP gap should be observed (1990, p.1328). The authors then proceed to run multiple experiments, each of which elicits WTA and WTP values for a particular market good and also for a 'token' which could be redeemed for a predefined amount money at the end of the experiment session. Any discrepancy between buying and selling prices caused by transaction costs, participant misunderstandings, or habitual strategies of bargaining, would be isolated by the WTA-WTP disparity, if any, observed for the tokens (Kahneman et al., 1990, p. 1328). If the WTA-WTP gap for the market good was greater than the WTA-WTP gap for the tokens, then this difference could be attributed to an endowment effect associated with the market good.

The results of Kahneman et al.'s study, found "the value that an individual assigns to objects such as mugs, pens, binoculars, and chocolate bars appears to increase substantially as soon as that individual is given the object" (1990, p.1342). Considering these objects all have easily accessible substitutes, and since the effects of the elicitation procedure itself were isolated by valuing tokens along with the market good, these findings provide strong evidence in support of the endowment effect theory and throw the alternative theories of weak experimental design and the substitution effect into question.

However, while Kahneman et al.'s findings did suggest that individuals experience some kind of an endowment effect, they were unable to provide a clear distinction between an endowment effect was caused by parting disutility and an endowment effect caused by unanticipated-ownership utility (1990). Research conducted by Loewenstein and Adler (1995) and Bischoff (2008) looked to make this distinction.

2.8.2 The Endowment Effect and Unanticipated-Ownership Utility

2.8.2.1 Loewenstein and Adler's First Experiment

Loewenstein and Adler (1995) suggest that the WTA-WTP disparity is caused by the fact that people get attached to goods that they own resulting in increased utility, but are unable to anticipate this feeling of attachment before the good is endowed to them. This theory aims to differentiate between parting-disutility and unanticipated-ownership utility, by collecting estimated selling prices from individuals prior to endowing them with a good. The essential "difference between the parting-disutility effect and ownership-utility effect is that the latter only explains the WTA–WTP gap if it is unanticipated, while the former provides an explanation without implying any bias in prediction" (Bischoff & Meckl, 2008, p. 1769).

To test their theory, Loewenstein and Adler construct two experiments which elicit selling prices before and after endowment, as well as buying prices. At the beginning of Loewenstein and Adler first experiment sessions, all subjects were shown a mug engraved with their university's logo - Carnegie Mellon and Pittsburgh University, depending on the location of the experiment group. A form was then randomly distributed to approximately half of all participants asking them to "imagine that they possessed the mug on display and to predict whether they would be willing to exchange the mug for various amounts of money" (Loewenstein & Adler, 1995, p. 931). They were given a list of 40 price levels ranging from 25cents to $10 at 25cent increments, and were asked whether they would rather keep the mug or trade it for the amount indicated on each line. After the subgroup of participants had completed their hypothetical choice tasks, all students were presented with a mug and given another set of choice tasks. The second set of instructions were identical to the first, but participants were told that one of the 40 selling prices had been pre-selected by the experimenter, and the choice made by each participant in regards to that line (to either to keep or sell the mug) would actually occur.

From this experiment, Loewenstein and Adler collected data from two groups of participants: those that had completed the prediction exercise prior to being endowed with the mug, and those that had not. This allowed the authors to

"conduct both a between - and within - subject analysis of prediction" (Loewenstein & Adler, 1995, p. 931).

Table 3. Predicted and Actual Valuation of the Mug

(Loewenstein & Adler, 1995, p. 932)

Group/condition	Number of subjects	Prediction of valuation	Actual valuation
Carnegie Mellon University			
Prediction	14	$3.73 (0.41)	$5.40 (0.65)
No prediction	13	—	$6.46 (0.54)
University of Pittsburgh			
Prediction	22	$3.27 (0.48)	$4.56 (0.59)
No prediction	17	—	$4.98 (0.53)

Std. errors in parentheses.

The results of Loewenstein and Adler's experiment, as seen in table 3, show that participants who took part in the prediction exercise significantly underestimated their own actual selling price by a difference of $1.67 at Carnegie Mellon ($p<0.02$), and $1.29 at the University of Pittsburgh ($p<0.01$). Those participants who did not complete the prediction exercise also stated actual selling prices well above the estimated selling price. This finding shows that participants placed a higher value on the mug once it was in their possession, and were unaware that this increase in value would occur prior to being endowed, suggesting an unanticipated ownership utility effect.

One limitation of Loewenstein and Adler's first experiment was that the elicitation procedure was not incentive compatible, so while respondents did not have any reason to misrepresent their true preference, there was no incentive to accurately reveal their true preference either (1995). A second limitation was that no WTP values were elicited from participants, so the authors were unable to determine how the predicted selling price (WTA_P) measured in relation to participants' actual WTP and WTA.

2.8.2.2 Loewenstein and Adler's Second Experiment

In their second experiment, Loewenstein and Adler (1995) aimed to solve the limitations of their first experiment by using an incentive-compatible elicitation mechanism, and by collecting WTP data from participants. To do this they randomly divided 106 students from the Northwestern University into two experiment groups based in two separate rooms.

In the first group a coin was tossed for each participant, and if they were able to correctly guess the outcome of the toss they were given a mug. Selling prices (WTA) were then elicited from those who had won a mug using the same procedure as in the first experiment, while buying prices (WTP) were elicited from those who had not won a mug, using a similar elicitation technique.

In the second group however, an identical mug was shown to participants, who were told that they had a 50% chance of winning a mug based on the result of a coin toss. Prior to the coin toss, predicted selling prices (WTA_P) were elicited from the entire group using the same payment-card style elicitation mechanism as in the first experiment. Participants were also told that one of their choices would be selected at random and would be the binding outcome if they were to win a mug. Next, the coin was tossed, and those participants who correctly called the outcome were given the good. If these 'owners' had agreed to sell the mug at a particular price level, predetermined by the experimenters but unknown to the participants, then they were now made to do so. Participants who were endowed with the mug were also then asked whether they would like to revise their selling price, although they were not actually allowed to do so. The results of this experiment can be seen in table 4.

Table 4. Mean Valuation of Mugs

(Loewenstein & Adler, 1995, p. 934)

Group	Form	Description	Number of subjects	Prediction of valuation
Control	1	Selling price	24	$5.96 (0.460)
	2	Choice	29	$4.05 (0.329)
Experimental	3	Selling price contingent on getting a mug	53	$4.16 (0.293)
	4	Desired revision of selling price	34	$4.69 (0.329)

Std. errors in parentheses.

The endowment effect is again evident in these results, with selling prices (WTA) for participants endowed with the good being significantly greater than the buying price of those participants not endowed with the good. More importantly however, is that a bias in prediction of selling prices is also evident. Participants "with a 50% chance of receiving a mug stated a mean selling price that was $1.80 lower than that for subjects who actually possessed a mug" (Loewenstein & Adler, 1995, p. 934). Furthermore, the selling price predicted by the un-endowed group was very similar to the buying price elicited from participants who had not won a mug in group 1 ($4.16 and $4.05 respectively). This indicates that subjects are unable to foresee the additional utility they would gain from possessing the mug, and therefore predict WTA values which were very close to elicited WTP values.

> *The observation that individuals are unaware of the endowment effect presents a novel view of choice. It suggests that people not only become attached to what they have (as implied by the endowment effect), but do so unknowingly. People seem to be unwittingly trapped by their choices; they make choices with an unrealistic sense of their reversibility*
>
> (Loewenstein & Adler, 1995, p. 936).

Loewenstein and Adler construct an index equation to measure the comparative influence that a parting disutility effect and an unanticipated ownership utility effect have on the WTA-WTP gap (1). If parting disutility was the sole cause of the WTA-WTP gap, "there is no reason to assume that the

members of the experimental groups should not be able to anticipate the disutility they would feel when having to part with the mug" (Bischoff, 2008, p. 286), and predicted WTA (WTA_P) should equal actual WTA, resulting in a beta value of zero in Loewenstein & Adler's equation (Loewenstein & Adler, 1995, p. 933). However, if 100% of the WTA-WTP gap is the result of an unanticipated ownership utility effect, then predicted WTA would equal actual WTP resulting in a beta value of one.

$$\beta = \frac{(WTA - WTA\_P)}{(WTA - WTP)} \qquad (1)$$

The prediction bias evident in the results from Loewenstein & Adler's second experiment, presented in table 4, is equal to 0.94 (2), which is 94% of its plausible maximum value (1995). This indicates that an ownership utility effect is causing a majority of the WTA-WTP disparity in the market for mugs, while a parting-disutility and other conventional factors plays a relatively insignificant role.

$$\beta = \frac{(5.96 - 4.16)}{(5.96 - 4.05)} = 0.94 \qquad (2)$$

Loewenstein and Adler make the comment that when stating a selling price for a good which the respondent does not possess, the individual requires two stages of introspection: "(1) imagining one possesses the object and has adapted to ownership, and (2) imagining how one would feel about parting with it" (1995, p.936). Based on the results of their two experiments, Loewenstein and Adler suggest that participants are not able to fully anticipate the additional utility that comes with the possession of the good, and therefore predicted selling prices are subject to a downwards bias. Stating a buying price however, only involves one stage of introspection, and Loewenstein and Adler state that they know of no evidence to suggest estimated WTP values would be biased (1995, p.936). In closing, the authors suggest areas of further investigation, noting that "it would be interesting to test whether people with objects overpredict the buying prices or choice values of those without such objects" (Loewenstein & Adler, 1995, p. 936).

An issue with Loewenstein & Adler's second experiment, is that the treatment group who predicted their own selling prices prior to being endowed with the good, knew that they had a 50% chance of winning the mug. It is therefore difficult to confirm that these participants did not feel in any way endowed with the good prior to estimating their WTA. If these students felt that they had a good chance of winning the mug, they may already consider the good to be part of their endowment, meaning an ownership utility effect may have introduced some bias into the predicted WTA values. This is one issue which other authors, particularly Bischoff (2008), have attempted to overcome.

2.8.2.3 Bischoff's Experiment

Bischoff's (2008) article *Endowment effect theory, prediction bias and publicly provided goods: an experimental study*, investigates the presence of a WTA-WTP gap for a public good, the connection between the observed disparity and the endowment effect, and the contribution of parting-disutility and unanticipated ownership utility to the size of the gap.

The author provides a brief summary of the endowment effect literature, concluding that WTA-WTP gap appears to be caused "either by the disutility from parting with one's endowment and/or by an extra utility from ownership", which may be unanticipated by non-owners (Bischoff, 2008, p. 283). According to Bischoff, the literature on the endowment effect has tended to focus mostly on privately owned goods (2008, p. 284), so his research aims to extend the literature by looking at the endowment effect in terms of publically owned goods. Bischoff's interest with investigating the existence of the endowment effect for publically owned goods comes from the fact that no one person holds exclusive property rights to such goods, so it is unclear whether individuals still experience 'ownership utility' for such a good. Furthermore, Bischoff aims to distinguish between an endowment effect caused by parting disutility, and an endowment effect caused by unanticipated ownership utility. Fundamental to Bischoff's experiment is the assumption that participants not endowed with a good who are asked to predict the WTA of participants who are endowed with the good, should

be able to account for the parting disutility effect experienced by the endowed group, but not the unanticipated ownership utility effect.

Bischoff's experiment was conducted as follows:

A class of finance students from the University of Giessen, Germany, were informed that they were being given the opportunity to take part in an exclusive tutorial session (the public good), currently unavailable to other students, which would help to prepare them for an upcoming finance exam. The students were told that the tutors were highly experienced and participation in the tutorial would be of great benefit. The class was then divided into 2 groups.

Group 1 was told that they were all entitled to attend the tutorial session free of charge, and were then asked to state (individually) their willingness to accept (the level of payment they would require to forgo attendance of the workshop). The group was instructed that if their average WTA was less that the per-person cost of running the tutorial ($C – decided on pre-experiment but unknown to participants until after) then no participants would be able to attend, and each person would receive the cost of the workshop in a personal payment ($C). However, if the group's average WTA was equal to or greater than 'C', then the original offer was kept and all students had the opportunity to attend the tutorial and no one would receive any kind of cash payment.

Participants in group 2 were each given an initial endowment of $15 and were then asked to state their willingness to pay for the opportunity to attend the tutorial session. If their average WTP was less than the '$C' (not yet known to the participants) then no one in group 2 would be offered the tutorial and all would retain their $15 endowment. If the average WTP was equal to or greater than 'C' then all participants would be able to attend the tutorial, and all would be required to pay the cost 'C' out of their $15 endowment regardless of whether or not they were planning on attending the tutorial.

The second part to this experiment was conducted 2 years after the initial WTA-WTP elicitation. This component involved a second class of finance students, also from Giessen University, who were chosen from a similar finance

lecture at roughly the same time in the year as the earlier trials. This ensured that the participants in the second experiment were similar, in terms of exam preparation as well as demographic characteristics, to the participants from the first experiment. This class was informed about the first experiment and the good (tutorial session) was described in a fashion identical to the previous experiment. The participants were then split into two groups: group 1a and group 2a.

Participants in group 1A were asked to give their best prediction of the average WTA of group 1 from the earlier experiment. As an incentive, payments were promised to those participants whose predictions were closest to the actual average WTA. Participants in group 2A were asked to give their best prediction of the average WTP of group 2 from the previous experiment. Again, incentives were used to increase the effort made by the participants. Group 2A is therefore expected to produce a WTA estimate for individuals who are not yet endowed with a good, and group 1A is expected to produce a WTP estimate for individuals who are not yet endowed with a good, as is standard with WTP experiments. Bischoff's analysis "is based on the assumption that the participants in group 2A are able to deliver an unbiased predictions for the WTP voiced in group 2" (2008, p. 291).

Table 5. Average WTP and WTA for the tutorial

(Bischoff, 2008, p. 292)

Group	Indicator	*N*	Average amount (standard error)
1	WTA	79	26.06 € (5.929)
2	WTP	78	8.88 € (0.483)
1A	Predicted WTA	103	21.86 € (2.496)
2A	Predicted WTP	59	8.17 € (0.456)

From Bischoff's results, presented in table 5, a WTA-WTP gap for the public good (the tutorial) can be clearly identified, producing a ratio of 2.93:1. The predicted WTP of group 2a (8.17) was not statistically different from the true WTP values elicited from group 2 (8.88), while the predicted WTA of group 1a (21.86) was significantly less than the real WTA (26.06) of group 1 at a 95% level

of confidence (Bischoff, 2008, p. 292). The predicted WTA was also significantly different from the true WTP. To calculate the share of the WTA-WTP disparity caused by unanticipated ownership utility, Bischoff uses Loewenstein and Adler's (1995) index formula (1):

$$\beta = \frac{(26.06 - 21.86)}{(26.06 - 8.88)} = 0.24 \qquad (3)$$

Bischoff concludes that the magnitude of the WTA-WTP ratio observed for the tutorial session is in line with what the relevant literature would expect. More importantly however, is that the "significant difference between predicted and actual WTA clearly supports the notion that the publicly provided tutorial is subject to an unanticipated ownership utility effect" (Bischoff, 2008, p. 293). The beta value of 0.24, from equation 3, indicates that the ownership utility effect accounts for roughly one quarter of the total WTA-WTP disparity, leaving three quarters of the gap explained by other conventional factors and possibly parting disutility.

To Summarise, Bischoff (2008) finds evidence that the endowment effect does apply to publically provided goods, with participants who were endowed with the tutorial session stating WTA values far exceeding the WTA values elicited from the non-endowed group. The author also finds that public goods are subject to an unanticipated ownership utility effect (24% of the WTA-WTP gap), even though the public good is not privately controlled, owned, or consumed by any one individual. The experiment also provides evidence that participants are able to produce unbiased estimates of other individual's WTP for a public good.

2.8.3 Substitutability and the Endowment Effect

It is important to note that while the two central explanations for the WTA-WTP gap (substitutability and the endowment effect) discuss the disparity from different theoretical perspectives, researchers accept that these two concepts are not mutually exclusive. Moreover, it is likely that the endowment effect is dependent on substitutability of the good being valued.

> *An owner will not be reluctant to sell an item at a given price if a perfect substitute is readily available at a lower price. This reasoning suggests that endowment effects will almost certainly occur when owners are faced with an opportunity to sell an item purchased for use that is not easily replaceable* (Kahneman et al., 1990, p. 1344)

Accordingly, the level of parting disutility experienced by an individual should be minimal if they know that there is a perfect substitute available. Furthermore, it may also be expected that the unanticipated ownership utility effect experienced when an individual is endowed with a perfectly substitutable good, would be less than the effect of being endowed with a rare good. However, there is little evidence to support this claim.

2.9 The Importance of understanding the WTA-WTP Gap

Understanding whether the WTA-WTP disparity is caused by the methodology itself or is an indication of an underlying change in preferences resulting from the ownership of a good is important because this knowledge can affect how decisions should be made regarding non-market goods. If the disparity is caused by weak experimental design, then it should be possible to control for these weaknesses and the benefits to society calculated from a buyer's (WTP) or seller's (WTA) perspective should more or less converge. However, if the gap is the result of an endowment effect, where owners of a good truly place a higher value on that good than non-owners, then the way property rights are established in CV experiments will have a significant impact on the way resources should be allocated.

Since holders of some 'right' (i.e. sellers) appear to value a good differently than non-holders, "one of the most economically consequential decisions [in experimental design] will be the initial establishment of the property rights, especially for environmental and other public amenities for which property rights are unclear" (Horowitz & McConnell, 2002, p. 428).

Horowitz and McConnell use the example of preserving land from development to show how the WTA-WTP gap could impact on how decisions are made (2002). The mean WTA/WTP ratio derived from their meta-analysis was 7, which would suggest "that the amount of land that would be preserved if development rights were held by the general public is 7 times higher than the amount that would be preserved if the rights were deeded to the landowner and had to be purchased by the public" (Horowitz & McConnell, 2002, p. 428). In other words, if the public was asked how much they would pay to preserve the land (i.e. purchase from the developers), they would be expected to state an amount 7 times less than they would charge the developers to purchase the land. Therefore, a measure of the 'correct' amount of land to preserve, based on the preferences of the public, would be highly affected by whether the elicitation question was phrased in terms of WTP or WTA.

Knetsch notes the if the underlying causes of the WTA-WTP gap "are not trivial ones attributable to wealth effects, but are those due to pervasive and large endowment effects... [then] economic assessments of losses will be seriously understated if willingness to pay measures are used" (1990, p. 230).

> *If environmental degradation or a contemplated change imposes losses on individuals, [the] usual practice of using the payment measure [WTP] will likely lead to large understatements of the welfare changes. As a consequence, too many environmentally disruptive projects will be encouraged, too many harmful activities will be allowed, inadequate mitigation measures will be undertaken when environmental values are at risk, and compensation for losses will not fully indemnify adverse welfare changes* (Knetsch, 1990)

Therefore, experiments which measure the WTA-WTP gap for different types of goods and services, and especially those articles which attempt extend the knowledge on the nature of the disparity are of significant importance in the non-market valuation literature. In order to accurately measure the true economic value of a non-market good, it is necessary to understand the process of how people value, and why people value goods differently.

2.10 **Literature Review Summary**

2.10.1 Non-Market Valuation

All decisions involve making a choice and all choices involve a sacrifice of some kind (Bateman et al., 2002, p. 2). When the true benefits of different courses of action are known, then the financial cost of each option can be weighed against the associated benefits to determine what decision will provide the greatest net gain. However, when a decision making task involves a non-market good or service, then no market-based price and quantity data is available with which to estimate the economic value of the good, and therefore the benefits of alternative courses of action are often unclear.

Non-market valuation is a "series of techniques that economists use to assess the economic value of market and non-market goods" based on the preferences of individuals (Lipton et al., 1995, p. 16). Through analysis of peoples' preferences towards different levels of provision of a non-market good, a demand curve can be derived and the total economic value that society places on that good can be estimated. Preference data can be collected either through revealed preference (RP) techniques, such as the travel cost method and the hedonic pricing method, or stated preference (SP) techniques, such as the contingent valuation method and choice experiments. Revealed preference methods use existing behavioural data to econometrically *reveal* the willingness to pay (WTP) of individuals for various attributes of a non-market good, while stated preference methods directly elicit valuation data from participants through decision making tasks.

The main benefit of RP techniques is that they use real behavioural data, and are therefore based on what people actually do rather than what they say they will do (Shaw, 2005), while their main disadvantage is that they cannot be used to value goods for which behavioural data does not exist. The main benefit of SP techniques is that the "models are capable of measuring a full range of values, including so called passive use or non-use values, as they do not rely on the observation of actual behaviour" (Du Preez et al., 2010, pp. 136-137). However, SP data may be subject to a wide range of biases resulting from the hypothetical nature of the decision making tasks, and weaknesses in experimental design.

2.10.2 Designing a Contingent Valuation Experiment

The contingent valuation method (CVM) is a SP technique which aims to measure the monetary value associated with the gain or loss of a specific non-market good by collecting willingness to pay (WTP) or willingness to accept (WTA) data from participants who are placed within a carefully designed hypothetical market scenario (Mitchell & Carson, 1989).

The hypothetical market model is the most critical component of a CV experiment because participants must make their valuation statements (WTP or WTA) contingent on the rules and changes outlined in the scenario (Bateman et al., 2002). This market scenario needs to describe the good being valued, the initial level of provision, the potential final level of provision (i.e. the gain or loss), the rule which dictates the final level of provision, any available substitutes, and the hypothetical payment mechanism (Mitchell & Carson, 1989, p. 3).

Once the market scenario is described to participants, they proceed to the decision making tasks which elicit their WTP or WTA values for the change in provision. While there are many different elicitation methods to choose from - open ended questions, the bidding game, payment cards, the single-bounded dichotomous choice method, and the double-bounded dichotomous choice method – each has its own advantages and disadvantages, with some techniques tending to create a bias in respondent's valuation statements (Bateman et al., 2002).

The final stage of creating a contingent valuation experiment is to collect data on participant's demographic characteristics, attitude and knowledge regarding the good being valued, and their views on the CV experiment as a whole. Demographic and attitudinal data can be used to assess the validity of an individual's valuation response, while participant's thoughts regarding the experiment can help determine if the CV design as a whole is likely to produce a reliable estimate of the population's WTP or WTA for a good. The validity of a CV experiment is based on the its ability to measure the true economic value that individuals place on a good, and is assessed in terms of *convergent validity* and *theoretical validity* (Mitchell & Carson, 1989). Reliability, on the other hand, refers to how much of the variation in WTP values (or WTA values) stated by respondents is caused by actual differences in the individuals' preferences, and

how much is caused by random sources, or 'noise', arising from the experimental process or biases in the sample group (Mitchell & Carson, 1989).

2.10.3 WTP and WTA Measures of Value

There are two types of valuation question which CV experiments can use to elicit preference data from participants. WTP questions ask respondents how much they are willing to pay to gain a good or to protect against losing a good, while WTA questions ask respondents how much they need to be paid to forgo the gain of a good or to accept the loss of a good (Mitchell & Carson, 1989, p. 25). Bateman et al. (2002) explain the concepts of WTP and WTA through the use of an individual's indifference curve, identifying four different measures of value: compensating variation for a gain, compensating variation for a loss, equivalent gain, and equivalent loss.

Compensating variation, either for a gain or for a loss, measures the change in private good expenditure (i.e. money), that would exactly compensate an individual for the proposed change in the non-market good in order to maintain their initial level of utility. The equivalent gain and equivalent loss measures of value represent the change in private good expenditure that would generate the same shift in the overall welfare level of the individual that the proposed change in the non-market good would create.

As shown previously in table 1, compensating variation for the gain of a good is the same as the equivalent loss for the loss of the good, which are both equal to the line BC, while compensating variation for the loss of a non-market good is the same as equivalent gain and is represented by the line DA. In Bateman et al.'s (2002) explanation of the different value measures, it is noted that WTA questions (line DA) are expected to result in slightly higher value estimates than WTP questions (line BC) due to an income effect experienced by those participants asked to state their WTA.

2.10.4 The WTP-WTA Disparity

According to standard economic theory, participants asked to state their WTA for a good experience a greater level of effective income than participants asked to state their WTP for the same good, and as a result WTA responses are

typically greater than WTP responses (Randall & Stoll, 1980; Willig, 1976). This disparity is caused because WTA respondents begin at a higher level of wealth than WTP respondents, holding all else constant, and 'wealthier' individuals are typically willing to pay more for an increase in a public good, or demand more for the loss of the good due to standard economic principal of the income effect (Bateman et al., 2002; Mankiw, 2007). Accordingly, WTA would be equal to the WTP if both measures were taken while holding the initial wealth level of participants constant. Many CV researchers point to the theoretical bounds for the WTA-WTP gap constructed by Willig (1976) and revised by Randall and Stoll (1980) which suggest that the divergence between these two value measures due to an income effect should be insignificant in most valuation experiments. However, "this near equivalence prediction has been contrary to the results of most empirical work trying to elicit WTP and WTA" which has regularly identified substantial differences between the two measures of value (W. L. Adamowicz et al., 1993, p. 417). Many researchers have therefore investigated alternative theories to explain the significant WTA-WTP gap observed in CV experiments.

2.10.5 Weak Survey Design

Some economists believe that the WTA-WTP gap is purely the result of imperfect experimental design where biases introduced through the survey process affect the way that participants interpret and respond to the WTA and WTP questions (Horowitz & McConnell, 2002). This theory suggests that neither WTA nor WTP represent the true preference of a participant, and that if all weak design features were accounted for, the two value measures would converge on the true value of the good. Three experimental design features which authors believe contribute to the WTA-WTP gap include: using a non-incentive compatible elicitation mechanism, failing to provide participant anonymity, and failing to provide clear instructions and training to participants regarding the elicitation mechanism.

An elicitation method is considered to be *incentive compatible* only if the technique provides "respondents with incentives to reveal their true valuation" of the good under analysis (Bateman et al., 2002, p. 128). Authors argue that the hypothetical nature of most CV experiments, where participants' decisions have

no real monetary consequence, are not incentive compatible, and that "the answers people give to hypothetical questions are likely to be off-the-cuff or careless responses which do not reflect their true taste preferences" (Carson, 1989, p. 172). A meta-analysis of 29 experimental found that on "average subjects overstate preferences by a factor of about 3 in hypothetical settings... [and that] willingness to pay studies yield smaller hypothetical-to-actual ratios than willingness to accept studies" (List & Gallet, 2001, pp. 241-251). This finding suggested that while both measures of value will be biased upwards in a hypothetical experiment, WTA estimates will be subject to a larger upwards bias than WTP estimates, therefore creating a large WTA-WTP disparity. A second meta-analysis conducted by Horowitz and McConnell (2002) however, found that "real experiments do not yield [WTA/WTP] ratios that are significantly different from those of hypothetical experiments" (2002, p. 437).

The second feature of the elicitation mechanism which is considered to create a large WTA-WTP gap is a non-incentive compatible provisional rule. Mitchell and Carson (1989) suggest that a respondent will only provide their true WTP (WTA) for a good if they believe that the provision of the good is contingent on their stated value and that they will be required to pay (or accept) the amount they state. If the participant believes that either their stated value will have no impact on the provision of the good, or that their stated value is disconnected from the actual amount they will be required to pay (or accept) for the good, then there is an incentive for strategic bias, and the value estimate will not reflect the respondent's true preference.

Failing to ensure subject anonymity is also argued to introduce a bias into the values elicited from participants. It is argued that participants making decisions without anonymity are more likely to think about how other people will view their decision and may therefore provide what they think is a 'socially acceptable' response, rather than a response based on their true preference. In a CV context, failing to provide anonymity may result in WTA estimates being overstated as respondents want to views as a "talented bargainer" (Plott & Zeiler, 2005, p. 538).

Allowing for participant learning is expected to increase how accurately WTP and WTA values capture a respondent's true preference towards a good, and

since these two measures should theoretically be identical (Willig, 1976), doing so is expected to eliminate the disparity. Several authors provide compelling evidence in support of this concept, finding that allowing for participant learning significantly reduces the WTA-WTP gap (Coursey et al., 1987; Shogren et al., 1994), but other investigations, such as Morrison (2002), found no such effect.

While Plott and Zeiler's (2005) experiment, which controlled for all of the features of weak experimental design detailed previously, found that WTA and WTP measures converged for a simple market good (a mug), evidence provided by Horowitz and McConnell (2002) and Kahneman et al. (1990), found that doing so did remove the disparity. Weak experimental design, while offering a partial explanation for the WTA-WTP gap, is generally considered not to be the sole cause of the disparity.

2.10.6 The Substitutability Effect

Hanemann's (1991) substitutability hypothesis explains the WTA-WTP gap as being caused by a diminishing marginal rate of substitution between private consumption and the non-market good being valued. Accordingly, goods that have easily accessible substitutes, such as pens and candy bars, should produce a lower WTA-WTP ratio than goods that do not have easily accessible substitutes, such as most public goods. While many experiments have produced findings in support of Hanemann's theory (W. L. Adamowicz et al., 1993; Horowitz & McConnell, 2002; Shogren et al., 1994), the authors also note the substitution hypothesis cannot account for the entire disparity between WTA and WTP, as a significant gap is still evident even for easily substitutable goods. Horowitz and McConnell (2002) for example, show that non-market valuation of ordinary private goods still produced an average WTA-WTP ratio of 2.92, when the substitutability hypothesis would predict a ratio closer to 1:1. Some researchers have therefore offered further theories regarding the WTA-WTP gap.

2.10.7 The Endowment Effect

An alternative explanation for the significant WTA-WTP gap is known as the *endowment effect*, and is based on the idea that a participant's true value for good increases once they are in possession of that good (Thaler, 1980). This argument suggests that participants asked to state their WTA hold property rights to the good where as WTP respondents do not, and that "this difference in point of reference causes a disparity between WTA and WTP" (Bischoff, 2008, p. 284). While following a similar line of thought to the conventional *income effect* explanation for the disparity, the endowment effect theory suggests that the gap is greater than the income effect would suggest due to either a parting disutility effect or an unanticipated ownership utility effect.

The parting disutility explanation for the endowment effect is based on Kahneman and Tversky's (1979) prospect theory which suggested that individuals place a higher value on the loss of a good than the gain of an identical good. Thaler suggested that "goods that are included in the individual's endowment will be more highly valued than those not held in the endowment, ceteris paribus,... because removing a good from the endowment creates a loss while adding the same good (to an endowment without it) generates a gain", and according to prospect theory this loss will be more highly valued than the gain (1980, p. 44).

The unanticipated ownership utility effect suggests that participants who are endowed with a good (i.e. the WTA respondents) become attached to that good resulting in a gain in welfare that they were unable to anticipate (Loewenstein & Adler, 1995; Bischoff, 2008)). If participants were able to fully predict this increase in utility then WTA and WTP should not differ (Bischoff & Meckl, 2008). However, if this extra utility is unanticipated, then only those participants endowed with the good experience the change which increases their WTA for the good.

It is also noted that the endowment effect would only create a WTA-WTP disparity if the good being valued did not have any perfect substitutes. If the good does have a perfect substitute then participants' indifference curves will be straight, and WTP will equal WTA.

Chapter 3 Method

This method section will begin with a review of the core research objectives (3.1) that this thesis aims to address. Following this is an overview of the experimental design, detailing the six different treatment groups that were required to the collect the data necessary to analyse these research objectives (3.2). Next, the construction of the six different survey designs (one for each of the six treatment groups) will be described and analysed (3.3), followed in section 3.4 by a description of the data collection process. Section 3.5 will address limitations of the experimental design, and section 3.6 will provide a brief summary of the methodology section.

3.1 Research Objectives

As outlined in the *Research Objectives* segment of this thesis in chapter 1, there are two core areas of investigation that this thesis aims to cover, both of which explore the underlying cause of the prominent WTA-WTP disparity commonly observed in CV experiments:

3) Investigation of the Endowment Effect

This research objective aims to isolate the effect that the endowment of a good has on participants' valuation of that good while controlling for other possible causes of valuation disparity, such as weak experimental design. Furthermore, this research aims to distinguish between an endowment effect caused by parting disutility and an endowment effect caused by unanticipated ownership utility.

To investigate the endowment effect explanation for the WTA-WTP gap, two research questions are posed:

e) Does a significant WTA-WTP disparity emerge in a CV experiment even when controlling for features of weak experimental design?
f) Are individuals who are not endowed with a good able to anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good?

4) Investigation of a Hypothetical Bias

This research objective aims to investigate how the hypothetical nature of a contingent valuation experiment impacts on the valuation estimates of participants, and also to assess whether a hypothetical bias contributes to the WTA-WTP disparity. To investigate these points, two more research questions are posed:

g) Does a purely hypothetical CV experiment produce significantly different valuation results to those obtained from an identical experiment but where participants' decisions have a binding monetary consequence?

h) Does a purely hypothetical CV experiment produce a significantly different WTA-WTP gap than an identical experiment with a binding monetary consequence?

In the following section, detail will be provided about the experimental design that was constructed to collect all of the data necessary to answer the research questions stated above. Once the design of the experiment is explained, and the various treatment groups have been identified, six hypotheses will be presented which relate to, and will be used to answer, the four research questions above.

3.2 Experimental Design Overview

This experiment adopts a design based on that used by Loewenstein and Adler (1995) and Bischoff (2008). Like these two previous research papers, one of the goals of this experiment is to isolate the effect that endowment has on an individual's valuation of a good, with the specific intention of separating the *unanticipated ownership effect* from the *parting disutility effect*. Therefore, following the methodology laid out in these previous investigations creates a reliable foundation for later analysis. This experiment also attempts to test whether the hypothetical nature of a contingent valuation experiment affects the WTP and WTA values stated by participants.

In order to achieve these goals, a laboratory-based contingent valuation survey was designed, which aimed to estimate the monetary value that participants place on a single attribute of a t-shirt, which was *certified organic cotton,* under differing experimental conditions.

3.2.1 The Good

Certified organic cotton was chosen to be the attribute under valuation because it holds certain 'public good' qualities, which were able to be emphasized through the design of the CV experiment. For example, certified organic cotton was defined as a *non-rival* good, since one participant receiving the good would not reduce any other participant's ability to receive the good. Organic cotton was also defined as a non-excludable good, since the outcome of the CV experiment determined whether the group as a whole received or did not receive the attribute, so no individual participant was able to be excluded from gaining the good once it was provided. See section 3.3.4 for more information on the rule of provision for the good.

The purpose of defining the organic cotton attribute as a public good was to try an replicate the experimental design used by Bischoff (2008), who used the CVM to value a tutorial session. Furthermore, many of the real-world applications of the CVM involve public goods, so basing this experiment on the valuation of a public good should increase the relevance of its findings.

3.2.2 What Data the Needed to be Collected

3.2.2.1 Binding WTP and WTA Treatment Groups

To identify the effects of endowment on participants' valuation of organic cotton, two treatment groups were needed: WTP_B, who were asked for their maximum willingness to pay to gain organic cotton, and WTA_B, who would be asked for their minimum willingness to accept to accept the loss of organic cotton. These two treatment groups were involved in a *binding* contingent valuation experiment, where the valuation decisions of the participants had real monetary consequences. This was to ensure the incentive-compatibility of the elicitation mechanism, and to allow for a comparison with value estimates collected from WTP and WTA groups where were involved in a purely hypothetical CV experiment.

Comparing the value estimates of participants not endowed with the good to the value estimate of those participants who are endowed with the good will provide a measure of how endowment affects an individual's valuation of organic cotton.

3.2.2.2 Prediction WTP and WTA Treatment Groups

To separate unanticipated ownership utility from parting disutility, estimated WTP and WTA data needed to be elicited from two more groups of participants, neither of which had been endowed with the organic cotton attribute. Following Bischoff's (2008) design, this required two *prediction* groups: WTP_P, who was asked to estimate the actual willingness to pay of WTP_B, and WTA_P, who was asked to estimate the actual willingness to accept of WTA_B.

According to the hypothesis of Loewenstein and Adler (1995) and Bischoff (2008), the WTP_P group should be able to provide an unbiased estimate of WTP_B group's valuation since they are at the same level of endowment (i.e. neither of them are endowed), while the WTA_P group should be able to provide an unbiased estimate of the WTA_B group's valuation only if there is no unanticipated ownership utility. Since the WTA_P group can not anticipate this

additional ownership utility, their WTA estimate should be less than the actual WTA of the binding treatment group.

How closely predicted WTA comes to actual WTA, assuming that predicted and actual WTP are equal, will provide a measure of the relative contribution of unanticipated ownership utility and parting disutility to the WTA-WTP disparity.

3.2.2.3 Hypothetical WTP and WTA Treatment Groups

To test whether the hypothetical nature of CV experiments is a contributing factor to the WTA-WTP gap, hypothetical WTP and WTA data is elicited under similar conditions to those set in the binding treatment groups. The hypothetical WTP group (WTP_H) underwent an identical procedure to the binding WTP group, except it was emphasized that there was no connection between participants' stated WTP values and the outcome of the experiment. The same conditions were set for the hypothetical WTA group (WTA_H).

By comparing the value estimates collected from the hypothetical treated groups to those elicited from the binding treatment groups, any *hypothetical bias* will be identifiable, and any contribution of such a bias to the WTA-WTP gap will be observed.

3.2.3 Summary of Treatment Groups and Construction of Hypotheses

As noted above, six different treatment groups are required to collect the data necessary for examination of this thesis' research objectives. These groups are:

WTP_B - Elicits WTP data from participants in a *binding* CV scenario

WTA_B - Elicits WTA data from participants in a *binding* CV scenario

WTP_H - Elicits WTP data from participants in a *hypothetical* CV scenario

WTA_H - Elicits WTA data from participants in a *hypothetical* CV scenario

WTP_P - Provides an estimate of the WTP_B data

WTA_P - Provides an estimate of the WTA_B data

Using data collected from these six treatment groups (above), four hypotheses can now be constructed which will help to answer the research questions detailed in section 3.1.

1) Does a significant WTA-WTP disparity emerge in a CV experiment even when controlling for features of weak experimental design?

$$H_0: WTA\_B = WTP\_B$$

$$H_1: WTA\_B > WTP\_B$$

Null: there is no significant difference between the value estimates collected from the WTA_B and WTP_B treatment groups.

2) Are individuals who are not endowed with a good able to anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good?

$$H_0: WTA\_B = WTA\_P$$

$$H_1: WTA\_B \neq WTA\_P$$

Null: there is no significant difference between the value estimates collected from the WTA_B treatment group, and those values predicted from the WTA_P treatment group.

3) Does a purely hypothetical CV experiment produce significantly different valuation results to those obtained from an identical experiment but where participants' decisions have a binding monetary consequence?

$$H_0: WTP\_B = WTP\_H$$

$$H_1: WTP\_B \neq WTP\_H$$

Null: there is no significant difference between the value estimates collected from the WTP_B and the WTP_H treatment groups.

H_0: WTA_B = WTA_H

H_1: WTA_B ≠ WTA_H

Null: there is no significant difference between the value estimates collected from the WTA_B and the WTA_H treatment groups.

3.2.4 Qualtrics – Online Survey Building Software

Since the experiments were conducted in the computer laboratories of the Waikato Management School, participants were able to be given web-based surveys rather than traditional paper-based surveys. This online data collection method had two key benefits: 1) it saved time and resources since the surveys did not need to be printed and distributed to respondents, and 2) since the data was already recorded in an online spreadsheet, time was also saved not having to input data from hard-copy surveys into a computer database.

Survey-building software Qualtrics was used (www.qualtrics.com) to construct the six different questionnaires required for the experiment. This software allowed for excellent customisation of the content and process of the survey, and offered several advantages over paper-based methods. For example, the bidding game mechanism which was used in the experiment was able to be customized so that once a participant selected the 'no-buy' option in the WTP groups or the 'exchange' option in the WTA groups (which represented their maximum WTP and minimum WTA respectively), the participant would skip to the next stage of the experiment, rather than going through the rest of the price levels which were irrelevant. The results of the Qualtrics-based survey were collected in a downloadable file which could be interpreted by Microsoft Excel, SPSS, and a number of other data-analysis packages.

The following section will provide a detailed look at the design of the six different surveys which were constructed using Qualtrics to collect the data necessary to test the hypotheses outlined in section 3.2.3.

3.3 **Construction of the Experimental Surveys**

As outlined in the *Experimental Design* overview (section 3.2), six separate surveys needed to be constructed to collect the different types of valuation data necessary for analysis. These six surveys, while differing slightly in design, all needed put participants in the same contingent valuation scenario, and elicit value measures from them through an identical process. This was to ensure that any difference in valuation results could be attributed directly to the intentional changes in survey design, rather than other random factors. The core differences between surveys lie in whether the survey was collecting **WTP** or **WTA** data and whether the experiment had **binding** monetary consequences, was purely **hypothetical**, or was a **prediction** exercise.

The following section will provide a detailed look at the design of the two contingent valuation surveys that were given to the **binding** WTP and WTA treatment groups, highlighting the small but critical differences between these two surveys. The examination will move through the survey process in the same way the participants did, and will address important design features such as: the design of the market scenario, the elicitation mechanism and provisional rule, and the follow-up questions that participants were asked.

Following this in-depth examination of the two *binding* CV surveys, brief detail will then be provided regarding the two *hypothetical* surveys (WTP_H and WTA_H) which are identical to their binding counterparts with the exception of a few changed in terminology used to emphasize the hypothetical nature of the experiment. Lastly, the surveys used to collect *predicted* WTP and WTA data will be addressed (WTP_P and WTA_P), and the incentive compatibility of the provisional rule will be examined.

3.3.1 The Binding WTP and WTA Surveys

3.3.1.1 Introduction - Defining the base-line level of provision

At the beginning of the survey, participants in the binding WTP and WTA groups were informed that have won a free t-shirt, and were shown an image of the t-shirt. Those individuals in the WTP group were informed that their t-shirt is made from non-organic cotton, while those in the WTA group were told that their t-shirt is made from 100% organic cotton. Participants were then asked to select their preferred t-shirt size and colour.

This introduction to the survey is a critical component to building the CV scenario because it defines the property rights for the attribute being valued, which in this case is *certified organic cotton*, and sets the initial levels of provision. Both groups were awarded with an identically styled t-shirts, but the provision of the organic cotton attribute is given only to those participants in the WTA treatment, and not those in the WTP treatment. This sets the two levels of endowment which is one of the focus points of this experiment.

Next, all participants were informed that the experiment was '*interested in [their] opinion about t-shirts made using certified organic cotton'*. Those individuals in the WTP group were told that their t-shirt can *'be made using 100% certified organic cotton, at an additional cost to [them], depending on the outcome of this decision making experiment'* (Appendix 1). Participants in the WTA group were told that the '*t-shirt that [they] have been given can be made using 100% certified organic cotton, at no additional cost to [them], depending on the outcome of this decision making experiment*',

The phasing of this information is again aimed to reinforce property rights for the organic cotton attribute while maintaining as much similarity as is possible between the information given to the WTP group and the WTA group.

3.3.1.2 Describing the good

According to Mitchell & Carson (1989) it is important to describe the good being valued so that participants are made aware of what exactly they are being asked to purchase or sell. It is possible that some participants may not understand what the term 'organic cotton' actually means, and it is likely that most will have limited knowledge of its benefits. Therefore, a brief paragraph on the definition and benefits of organic cotton was offered to participants.

This information described how cotton is 'traditionally known as the world's 'dirtiest' crop because of the large amount of synthetic fertilizers and insecticides used in its production', and how 'certified organic cotton is grown using methods and materials that have a low impact on the environment' (see Appendix 1**)**. It is also noted that organic cotton production systems can help to replenish and maintain soil fertility, reduce the use of toxic and persistent pesticides and fertilizers, and help to build biologically diverse agriculture

This information aimed to help participants become more familiar with the concept of organic cotton, and the benefits associated with organic cotton production.

3.3.1.3 Instructions - Elicitation mechanism and provisional Rule

Once the good (organic cotton) had been described to participants, they were then provided with instructions which introduced to the elicitation mechanism and also the rule which would decide the final provision of the good.

WTP_B Elicitation Mechanism and Provisional Rule

Respondents in the WTP_B group were informed that a '*number of [their] fellow participants [had] also just been given a free non-organic cotton t-shirt, and a set of instructions identical to [theirs]'* (Appendix 2)

Participants were then informed they were going to be given a *'series of options where [they] will indicate whether [they] would prefer to keep the non-organic cotton t-shirt or exchange it for an identical certified organic cotton t-*

shirt for which [they] would be required to pay the amount indicated'. An example of the elicitation question was then provided to them, and they were told that the price level would increase at $1 increments until they were no longer willing to pay for the organic cotton upgrade, and this would represent their maximum willingness to pay.

The provisional rule was then introduced to the binding WTP group, a rule which was based on whether the average maximum WTP of the group as a whole was greater than the actual cost associated with the organic cotton attribute, denoted by $C. This *cost of organic cotton* ($C), was set at $4 by the experimenter, but was unknown to the participants. This cost estimate for organic cotton was based on observations from a web-based wholesalers which sold both organic and non organic cotton t-shirts, but the actual level of $C is largely irrelevant to the experimental process.

Participants in the WTP group were then told that if the *"**average** [maximum] amount the group is willing to pay is **equal to or greater than** $C, then all of [their] t-shirts will be made using organic cotton and [they] will each be required to pay this additional cost ($C) before [they] are able to pick up [their] t-shirt*' (Appendix 2). They were also informed that if the average maximum WTP of the group was less than $C, then no one would receive the organic cotton attribute, and no one would be required to make any payment. This provisional rule is theoretically incentive compatible, meaning it is in the respondent's best interest to answer truthfully, for the reasons outline in section 3.3.3.

WTA_B Elicitation Mechanism and Provisional Rule

Respondents in the WTA_B group were informed that a '*number of [their] fellow participants [had] also just been given a free 100% Certified Organic Cotton t-shirt, and a set of instructions identical to [theirs]'* (Appendix 3).

Participants were then informed they were going to be given '*a series of options where [they] will indicate whether [they] would prefer to keep the*

certified organic cotton t-shirt or exchange it for an identical non-organic cotton t-shirt for which [they] would receive a personal payment of the amount indicated'. An example of the elicitation question was then provided to them, and they were told that the price offered would increase at $1 increments until they were willing to exchange their organic cotton shirt for the non-organic shirt.

The provisional rule was then introduced to the WTA group, a rule based on whether the average minimum WTA of the participants was greater than the actual cost associated with the organic cotton attribute, denoted by $C.

Participants in the WTA group were then told that if the '***average*** *minimum payment demanded by the group is* ***less*** *than the $C, then all of [their] shirts will be made from non-organic cotton and [they] will each receive a payment of $C, regardless of whether or not this is [their] personal preference*' (Appendix 3**).** Participants are also informed that if the group's average WTA is equal to or greater than $C, then they all retain their organic cotton t-shirt, and no one receives any further payment.

3.3.1.4 The Value Elicitation Process

Once the elicitation mechanism and the provisional rule had been introduced, participants moved on to the real elicitation choice questions. Before they began their first choice task, the binding nature of their decisions was again reinforced.

An iterative *bidding game* was chosen to be the elicitation mechanism used in this experiment because its simplicity makes it easy for participants to understand, and because it aims to directly capture the maximum (minimum) amount an individual is willing to pay (willing to accept) for a good (see section 2.23). Alternative mechanism, such as the dichotomous choice approach, require the use of econometric regression techniques to derive participants' maximum WTP or minimum WTA, and often require a much larger sample size to produce value estimates of statistical significance (Mitchell & Carson, 1989, p. 101). Using such mechanisms would add unnecessary complexity to the data-analyse

process, and would reduce the statistical power of the value estimates derived from the relatively small participant sample.

The *willingness to pay* elicitation questions asked participants to select their most preferred of two options: a) *'Keep non-organic cotton t-shirt, or'* b) *'Exchange for an identical* ***organic cotton*** *t-shirt and pay $_X_'*, where the value of X increased from $1 to $20 at $1 increments (Appendix 4). If a participant selected option-b, indicating that they were willing to pay the amount X to purchase the organic cotton attribute for t-shirt, then the price would increase by $1 and they would again be asked to choose their most preferred option. This process would continue until either the participant selected the 'Keep non-organic cotton t-shirt' option, or they reached the end of the payment ladder, at which point they would move on to the next stage of the survey. The highest $X amount that an individual was willing to pay to upgrade their non-organic cotton t-shirt to an organic cotton t-shirt is interpreted as their maximum WTP for organic cotton.

The *willingness to accept* elicitation question however, asked participants to select their most preferred of two options: a) '*Keep organic cotton t-shirt, or'* b) '*Exchange for an identical non-organic cotton t-shirt and $_X_'*, where the value of X again increased from $1 to $20 at $1 increments (Appendix 5). If a participant selected option-a, indicating they were not willing to sell the organic cotton attribute of their t-shirt for the amount X, then the price would increase by $1 and they would again be asked to choose their most preferred option. This process would continue until the participant either selected the sell-option (option-b), or reached the end of the payment ladder, at which point they would move on to the next part of the survey. The point at which a participant switched from option-a to option-b is considered the minimum compensation they require to accept the loss of the organic cotton attribute (i.e. WTA).

3.3.1.5 The Prediction Exercise

After the maximum WTP or minimum WTA had been elicited from the participants in the two binding groups, they were then presented with a prediction exercise where they were asked to estimate the average valuation collected from the other group. This meant that the WTP_B group was asked to estimate the

average WTA of the WTA_B group, and vice versa. Further detail regarding the process of this prediction exercise, as well as the four other prediction exercises included in this experiment, will be provided in section 3.3.3.

3.3.1.6 Follow-Up Questions

Once participants in the two binding groups had completed their respective elicitation and prediction exercises, they were presented with a series of 'follow-up' questions. This section was identical for *all* participants in all six of the treatment groups,

The follow-up questions collected standard demographic data from the participants including their: age, gender, occupation, weekly disposable income, number of dependents, education, and ethnicity. This demographic information would be regressed against participants' WTA and WTP to ensure their valuation responses were not purely random (Mitchell & Carson, 1989).

Following these questions, participants were thanked for their time, and told to see the experiment supervisor to collect their $25 compensation payment. They were informed that their t-shirt, made from either organic or non-organic cotton, depending on the outcome of the experiment, would be ready to be picked up within two of three weeks, and that they would need to exchange their participant card for their t-shirt (since no names were collected during the experiment). This ended the binding experimental procedure.

3.3.2 The Hypothetical WTP and WTA Surveys

The two hypothetical CV surveys (WTP_H and WTA_H) were identical to their binding counterparts, except that various words had been changed to ensure respondent's understood that the experiment was purely hypothetical. The information on the benefits of organic cotton, the initial setting of property rights, the instructions that were given, and the elicitation mechanism and provisional rule that were used were all identical to those previously described in section 3.3.1, with the exception of several differences in terminology. These differences will now be addressed with regards to the WTP treatment only. The WTA_H

survey is identical to the WTA_B survey, except for the same changes in terminology that will now be identified for the hypothetical WTP survey.

In the beginning of the experiment, when the property rights were being defined, participants in the hypothetical WTP group were told to '***Imagine*** *that [they had] just won a free t-shirt made from Non-Organic*, as oppose to the introduction of the WTP_B group which informed participants that they '*had'* won a real t-shirt. Furthermore, in the instructions section of the survey, the word '*imagine'* was added in some places and phrases such as '*would be willing to pay'* were replaced with '*would have been willing to pay'*, in order to emphasize the hypothetical nature of the experiment. Prior to beginning the elicitation procedure, participants in the hypothetical treatment groups were asked to '*remember that the results of this experiment are* ***purely hypothetical****, but try to answer as though the scenario is real'*

The wording of the elicitation questions for the two hypothetical groups was the same as for the binding groups, as was the prediction exercise and the follow-up questions. More detail on the prediction exercise will be presented in the following section (3.4.3).

3.3.3 **The Predicted WTP and WTA Surveys**

The two surveys that were used to collect **estimated** WTP and WTA data (WTP_P and WTA_P) aimed to place respondents within a CV scenario identical to the one used to elicit the actual **binding** WTP and WTA data. According to Bischoff, "when having to predict the preference of others, subjects anchor on their own preferences", so the closer the preferences of the prediction groups' matched those of the binding groups', the more closely predicted values should match actual values. It was therefore critical to ensure that factors which may influence participants' preferences, such as demographic characteristics and the survey process, were identical in both the prediction and binding groups.

Random allocation of participant into the six different treatment groups ensured that demographic characteristics should be similar across all groups.

Therefore preferences due to the age, gender, ethnicity, occupation, and education of participants should also be equal across all groups. To further ensure that the preferences of the prediction groups were similar to those of the binding groups, the two prediction surveys were identical to their binding counterparts except that it was made clear to participants that they were to try and **predict** the average valuation made by the binding groups. The prediction surveys will now be described with reference to the WTP_P survey only, which ran in an identical fashion to the WTP_B survey, with a few small changes. The WTA_P survey is identical to the WTA_B survey, with the same necessary alteration which will now be outlined below for the WTP_P survey.

The WTP_P survey began by informing respondents that '*A number of [their] fellow participants, selected at random, [had] each won a plain t-shirt made from non-organic cotton, similar to the one seen below* [an image of a t-shirt was inserted]'. The prediction participants were able to customise the t-shirt in the same way the binding group had done so. Next, respondents were informed that this group (WTP_B) had been given the option to exchange their non-organic t-shirts for identical organic cotton t-shirts, depending the group's average willingness to pay for the upgrade. The prediction participants were told that they were going to be given the same survey that the WTP_B group had received, but that they should '*answer the survey questions on the bases of how you believe they [the WTP_B respondents] will respond to them'* (Appendix 6). An incentive to encourage thoughtful prediction estimates was introduced by offering a $40 cash reward to the participant able to provide the most accurate prediction of the average WTP of the binding participants.

Following this introductory information, the survey proceeded in an identical fashion to that of the WTP_B survey, explaining the benefits of organic cotton, the elicitation mechanism, and the provisional rule that would be used. Before they began the actual elicitation procedure however, they were told to '*Remember that [they] are answering these questions on the basis of how [they] believe [their] fellow participants will answer them, noting that their [the binding group] decisions have real monetary implications'*. The $40 monetary reward was

also reinforced, before participants proceeded with the bidding game elicitation mechanism.

Once they had completed the prediction elicitation process, the WTP_P group were given the same follow-up questions that all respondents received, and were then asked to leave the experiment room. The differences between the WTP_P and WTP_B groups detailed above were identical to the differences between the WTA_P and WTA_B groups.

3.3.3.1 Additional Prediction Surveys

As well as the two primary prediction exercises (WTP_P and WTA_P), prediction data was also collected from the four other treatment groups after they had each completed their main elicitation task. The binding and hypothetical WTP groups were both asked to estimate the average WTA of the WTA_B participants, while the binding and hypothetical WTA groups were both asked to predict the average WTP of the WTP_B participants. These additional prediction exercises were conducted using the exact same survey design that was given to the two main prediction groups.

3.3.4 Incentive Compatibility of the Provisional Rule

The provisional rule used in this experiment, described in section 3.3.1, was designed to be incentive compatible, meaning that it is in participant's best interest to state their true WTP or WTA for the good. Its incentive compatibility is reliant on the fact that participants do not know the value of $C, which removes the strategic incentive to overstate or understate their true valuation of the good to try and influence the average WTP or WTA of the group. To demonstrate this point, the thought process of a respondent stating their WTP when $C is known and when $C is unknown will be examined.

If a participant's true WTP is $10 and the value of $C is known to be $7, then the participant has an incentive to overstate their true valuation to increase the average WTP of the group and ensure provision of the good, because they know they will only be required to pay the $7. If a participant's true WTP is $5

and $C is known to be $7, then there is an incentive to understate their true WTP to reduce the average WTP of the group, and have the good withheld.

This incentive to produce biased valuation data results from the fact that participants believe that their stated WTP will influence the provision of the good, since provision is based on the group's average valuation, but is unrelated to the actual amount they will be required to pay, since this amount is determined by $C (Bateman et al., 2002, p. 128). To remove this incentive for strategic bias, $C is therefore made unknown to participants, thereby creating a connection between stated WTP and the actual amount participants will be required to pay.

If a participant's true WTP is $10 and the value of $C is unknown, then an overstated WTP may result in the participant actually having to pay more than $10 if the good is provided (e.g. is $C was $12). An understatement of true WTP may result in the good not being provided even thought the participant would have been willing to pay the amount necessary to obtain the good (e.g. if $C was $7). The fact that $C is unknown to the participant means they should not understate or overstate their true WTP or WTA, because doing so may result in an unfavourable outcome for the participant. Therefore, the provisional rule outlined above is considered incentive compatible.

In the following section, the process of data collection will be examined, looking firstly at how participants were recruited, and secondly at the actual experimental process.

3.4 Data Collection

3.4.1 Collecting Participants

To gather participants for the experiment, a number of different recruitment strategies were utilised, which included: talking directly to students during a summer school class, starting a number of email-chain to spread information about the experiment, distributing flyers throughout the Waikato University campus, and placing a job advertisement on a student employment website. In order to encourage involvement, a $25 participation fee was promised to each respondent to compensate them for their time.

On the first of December 2010, a third-year finance class of around 40 students was given a brief presentation which explained the details of the experiment, the compensation participants would receive, and instructions on how to register for one of the six experiment sessions. Students were free to ask further questions regarding the experiment to clarity any points.

Due to ethical restrictions on the use of Waikato University's student email distribution list, it was not possible to publicise the experiment to all students through a single bulk email. A chain-email approach was instead adopted where students were encouraged to distribute a information sheet about the experiment to friends that they believed might be interested in taking part in the experiment. This information sheet (Appendix 7) provided brief details about the experiment, such as how long the session would take, what the research was investigating, and the compensation payment participants would receive for their time. Copies of this experiment information sheet were also posted on 12 notice boards around the University campus, particularly in areas of high foot traffic so as to maximise exposure.

As well as the in-class presentation, the chain-email system, and the distribution of flyers around the university, a job-placement was also advertised on the Student Job Search website (www.sjs.co.nz). This advertisement stated that student participants were wanted for an economics experiment that would be held in the Waikato Management School computer labs, and that participants would receive a payment of $25 for roughly one hour of their time.

The response to the various forms of recruitment, particularly the SJS notice, was substantial, and within two weeks of the web-based registration

system being opened, the six experiment sessions were fully booked with 180 registered participants. The number of participants allowed to register for each experiment session was limited to 30 to make conducting the sessions easier for the experimenters, and also due to limitations on the number of computers available in the laboratory.

3.4.2 Collecting Data

Data was collected during six experiment sessions which ran from the 13th of December 2010 till the 16th of December 2010 in the computer labs of the Waikato University Management School. Each session involved between 27 and 30 participants and lasted up to one hour, with the first half of each session dedicated to a contingent ranking experiment being conducted by Professor Riccardo Scarpa, and the second half dedicated to the contingent valuation experiment detailed previously.

The decision to run both the contingent ranking experiment and the contingent valuation experiment consecutively was based on both financial constraints and time constraints. Running the experiments independently would have meant additional participants would have needed to be recruited and further compensation payments would have needed to be made to ensure the desired quantity of respondents were attained for both experiment. Joining the two experiments together was therefore convenient, and since no adverse impacts on either the ranking exercise or the CV exercise were expected, the decision would have a minimal effect the results of the experiments.

As previously noted, there were six experimental treatment groups involved in the CV experiment, each with a different survey design: WTP_B, WTA_B, WTP_H, WTA_H, WTP_P, and WTA_P. During each of the six experiment sessions, participants were divided evenly and randomly into these six treatment groups through the use of randomly-distributed participant cards, and data was collected accordingly. This process had the advantage of ensuring a random allocation of participants to each of the six treatment groups. If each experimental session has been dedicated to just one the six treatment groups, then the sample for each treatment may not have been random since the day and time of each experiment session may influence the type of participants involved (e.g. a

higher proportion of people with jobs may participate in the 6pm - 7pm sessions due to work obligations).

3.4.3 **The Experimental Procedure**

The procedure for each of the six experiment session was identical to ensure that no bias was introduced into the results from differences in the experimental process. Each session lasted up to one hour, with half of that time allocated to Professor Scarpa's contingent ranking experiment and the remainder allocated to the CV experiment. What follows is an overview of how a typical experiment session was conducted.

At the beginning of the experiment session each person randomly selected a 'participant card' which provided them with a unique participant-id number, a treatment code for the experiment, and a unique security code. Participant-id numbers ranged from 100 to 130 for the first experiment, 200 to 230 for the second experiment, and so on, and were the only form of identification attached to respondents' answers. Treatment codes ranged from one to six (i.e. for each of the six subgroups) with roughly five of the thirty participants being allocated to each of the six treatment codes in each experiment session. See Appendix 8 for an example of the participant card design. Participants were informed that these cards were important and that they should keep them in a safe place after the experiment, since any prizes won during the session, such as t-shirts or prediction winnings, would need to be collected upon redemption of the participant card.

Once each person had received a participant card they were asked to find an unoccupied computer and read through the information sheet which was displayed on the screen, ticking the 'I agree' box if they accepted the conditions specified on the form. See Appendix 9 for the *Information Sheet for Participants*. Participants were asked not to begin the survey until they had been given further instruction.

After a majority of the thirty participants who had registered for the session had arrived and taken their seats, the *introductory dialogue* (Appendix 10) was read aloud to the class. This dialogue introduced the two experimenters, Riccardo Scarpa and Francis Powley, and clarified the structure that the experiment session would take. Participants were informed that the unique id

number they had each been given would be the only connection between themselves and the responses that they provide, ensuring their complete anonymity. They were asked to read all of the information in the survey thoroughly and to pay particular attention to the instructions they were given. They were told that after completing their first set of ranking choice tasks they would need to enter their participant-id and treatment-id numbers before continuing to the second half of the experiment (the CV component). Participants were advised to raise their hand if they had any questions or ran into any technical difficulties during the session in order to alert the experimenters. It was also noted that the experiment was investigating their individual preferences towards the each of the options they were to be given, so there were no *right* or *wrong* answers. Lastly, they were told that they could collect their $25 compensation payment outside the class room once they had completed the experiment.

Once the introductory dialogue was complete, participants were again asked to thoroughly read through the *Information Sheet for Participant*, if they had not already done so, and were then told they could begin the experiment.

During the subsequent hour, participants each worked through their survey individually, with only a few technical difficulties arising, most of which related to the images failing to load on the computer screen. These issues were quickly resolved with only a minor interference into the experimental process. As participants each completed the experiment, they left the room to collect their $25 and were asked again to keep their participant card safe if they wished to collect any prizes they may have won.

The six experimental sessions were all conducted in the way described above, with no distinguishable difference in process between the sessions. Therefore, participants within each of the six treatment groups received identical information and underwent an identical survey procedure even though they were divided over six different session times. Furthermore, a majority of the information which participants received was provided via the computerised survey itself, which again helped to minimise any difference in procedure between sessions. Ensuring that all participants underwent the same experimental procedure was vital since the data collected from a particular treatment group (e.g.

WTP_B) in a single session (e.g. session 1) would be aggregated with the associated data from the five other sessions, and any differences in the data collection process would make this aggregation problematic.

3.4.4 Ethical Issues

To comply with the ethical research guidelines put in place by the Waikato University, it was necessary to provide participants with full anonymity. To achieve this goal each person was given a unique *participant card* upon entry into the computer labs which provided a *respondent number* as well as a *treatment code*. The individual was asked to enter these details into the survey system prior to commencing the experiment.

Other than anonymity, there was no ethical concerns regarding the contingent valuation experiment, and the research was granted ethical approval by the Waikato Management School Ethics Committee. See Appendix 11 for the accepted ethical application form.

3.5 Limitations

3.5.1 Theoretical Validity

The validity of a CV study is a related to how accurately the experiment is able to measure the true economic value that individuals place on a good. The concept of validity is separated into convergent validity, the degree to which an alternative non-market valuation technique produces WTP estimates which converge with those of the CV experiment, and theoretical validity, which is how closely the experiment's findings are consistent with economic theory (see section 2.3.3 - Evaluating the Validity and Reliability of CV Data).

Testing for theoretical validity usually involves regressing an individual's WTP against independent variable which hold some theoretical connection to their WTP, such as their attitudes towards the good (Mitchell & Carson, 1989). However, the experiment presented previously did not collect data on participants' attitudes towards organic cotton or the benefits associated with organic cotton production, so testing the theoretical validity of the valuation estimates is problematic. Without attitude data from participants a comprehensive test for theoretical validity is not achievable, though it may be possible to regress the WTP of participants on their demographic characteristics in order to make some loose inferences about the validity of the experiment results.

Failing to demonstrate the validity of the CV results would make it difficult to justify the accuracy that that value measures collected from the participants represents their true preferences. However, since the goals of this research are to test for differences in valuation estimates coming from hypothetical and binding scenarios, and from WTP and WTA elicitation mechanisms, validity is not a large concern.

3.5.2 Participant Learning

The experimental design failed to account for the effects that *participant learning* may have on the magnitude of the WTA-WTP gap. Several authors had found evidence that running a number of practice elicitation rounds before eliciting real value estimates from participants reduced the size of the disparity as respondents became more familiar with the elicitation procedure and were able to more accurately indicate their true preferences (Coursey et al., 1987; Shogren et

al., 1994). While the experimental evidence on the effects of participant learning are still mixed (Morrison, 2000), controlling for this feature of weak experimental design would give more credibility to the findings of this research.

However, an example of an elicitation question was shown to participants before they began the elicitation procedure (Appendix 4 and 5**)**, and they were also able to move backwards through the survey to change their responses if they wished. While not an ideal way to facilitate participants learning, these design features would still have removed some of the confusion relating to the elicitation procedure.

3.5.3 Sample Biases

Contingent valuation researchers typically define the population they want to estimate WTP or WTA values for (i.e. the population that would be affected by a proposed policy change), and then collect a random sample of participants from within this population (Carson, 1989). Collecting a random representative sample allows the experiment's valuation results to be generalised to the greater population from which the sample was taken.

However, in the CV exercise outlined previously, no experimental population was defined, and random sampling techniques were not used to collect respondents (see section 3.4.1 – Collecting Participants), and therefore a sample selection bias may be present within the participant group. For example, many of the participants were recruited through the Student Job Search advertisement, so the sample may have a high proportion of unemployed or student respondents, who may be less willing to pay for organic cotton, then the general Hamilton population. It would therefore be inappropriate to suggest that WTP of the sample group could be used to estimate the WTP of the Hamilton population.

For the purposes of this research, sample selection bias is not an issue provided participants were divided into each of the six treatment groups (WTP_B, WTA_B, ect.) randomly, which they were. Consequently, any biases in the experimental sample would be identical in each of the treatment group, so while the valuation results themselves may not be generalisable, comparisons of WTP and WTA data collected from each group would still provide robust findings.

3.6 Summary: Methodological Design

3.6.1 Identifying the Endowment Effect

The two surveys used to test for the affect that endowment has on participant's valuation of a good, WTP_B and WTA_B, were identical except for the small changes required to set the two different levels of endowment, and to collect the 'buying' and 'selling' prices. In order to isolate the *endowment effect* explanation of the WTA-WTP gap from other theories such as weak experimental design, it was critical that the binding surveys were constructed in such a way to reduce other possible causes of bias.

To account for these other potential causes of the WTA-WTP disparity, the experimental design utilised an incentive compatible provisional rule – so there is no incentive for strategic response behaviour, the decision-making tasks had binding monetary outcomes - so participants should make thoughtful choices, participant anonymity was guaranteed - so participants are able to express their true preferences without the influence of social pressures, and although there were not practice elicitation rounds, the mechanism was clearly explained – so participants should be able to present their true valuations trough the bidding game mechanism.

By controlling for these weak experimental design factors, any disparity between WTA_B and WTP_B can be attributed to the different levels of initial endowment experienced by the two groups: the un-endowed WTP group, and the endowed WTA group.

3.6.2 Isolating the Unanticipated Ownership Utility Effect

As previously noted, the process used to elicit the predicted WTP and WTA data was identical to the method used to elicit the binding WTP and WTA data, except that participants were asked to answer the valuation questions how they *believed* their binding counterpart groups would respond to them. As noted by Bischoff (2008), there is scarce experimental evidence to support the notion that predicted valuation estimates will converge with actual valuation estimates in a CV experiment. However, if predicted WTP is able to offer an unbiased estimate of actual WTP, then the suitability of the prediction methodology used in this experiment is establish. Then, if predicted WTA diverges from actual WTA,

as was found by Bischoff (2008) and Loewenstein and Adler (1995), then this is an indication that an unanticipated ownership utility effect is present in the valuation data. If no divergence between predicted and binding WTA is observed, then this contradicts the argument of an unanticipated ownership utility effect, and suggests that the WTA-WTP gap is more likely the result of people's aversion to loss (i.e. the parting disutility effect).

3.6.3 Isolating a Hypothetical Bias

Since the two hypothetical treatment groups are identical to their binding treatment counterparts, except for the few changes necessary to emphasize that they are purely *hypothetical*, any differences in valuation results between the two sets of data can be attributed directly to a hypothetical bias. Furthermore, if there are any unanticipated imperfections in the survey methodology, the bias in WTP and WTA data resulting from these flaws would be the same in both the binding and hypothetical treatments. Therefore, comparing the data collected from the WTP_B and the WTP_H groups, as well as the WTA_B and the WTA_H groups, will allow for a clear identification of the effects that the hypothetical nature of contingent valuation experiments has on participants' stated preferences.

In the following section of this thesis, the findings of the experiment, based on the data collected using the six survey designs outlined previously, will be presented along with discussion on these findings.

Chapter 4 Findings and Discussion

This chapter will present the valuation data collected from the six survey groups detailed in section 3.2.3 along with findings and discussion relating to each of the research objectives outlined in section 3.1 This chapter will begin by presenting the summary statistics derived from the 10 sets of data collected from the 6 treatment groups (section 4.1). This data will be tested to: 1) ensure that there was no significant difference between the demographic characteristics of the sample groups which could have introduce a sample selection bias, 2) assess the data's reliability and theoretical validity, and 3) determine whether the data sets follow approximate to a normal distribution curve (section 4.1). Section 4.2 will assess the existence and magnitude of a WTA-WTP gap using the binding treatment data. Next, the predicted WTP and WTA data will be used in conjunction with their binding equivalents to indentify whether WTA_B respondents experience an unanticipated ownership utility effect, and how this effect contributes to the WTA-WTP gap (section 4.3). Section 4.4 will compare the hypothetical WTP and WTA data to their binding equivalents to assess whether the hypothetical nature of CV experiments have a significant impact on the stated values of participants. Section 4.5 will then review the predicted WTP and predicted WTA data elicited from the hypothetical and binding groups, to determine how the elicitation process prior to the prediction exercise affects the ability of participants to provide unbiased estimates other participant's stated values. Lastly, a summary of the finding and discussion will be presented (section 4.6).

4.1 Assessing the Quality of the Data

This section will firstly present the summary statistics derived from the WTP and WTA data sets that were collected through the CV experiment detailed previously (chapter 3), and some brief comments on these statistics will be made.

Next we will assess whether any fundamental differences exist between the demographic characteristics of the six different treatment groups. The hypothesis testing conducted in later sections of this chapter relies on the assumption that any differences in valuation estimates are the direct result of the specific changes made to the CV surveys. Any difference between treatment groups, other than the intentional changes in the survey design, could introduce a variety of other participant-based biases into the results.

Following this, the reliability of the value estimates elicited from each of the six treatment groups will be tested by regressing the WTP or WTA on the demographic characteristics of the participants in that group. This will determine whether variation in the WTP or WTA responses is the result of underlying differences in participant's valuation of the good, or simply random 'noise' arising from the survey process (Mitchell & Carson, 1989, p. 211).

Next, the WTP data from the binding treatment group will be used to estimate a demand curve for organic cotton, and to calculate the welfare gain associated with provision of the good. This will demonstrate what a real-world application of the CVM may aim to calculate. However, considering this is not the focus of the research, welfare gains or losses will not be addressed in great detail.

Lastly, the WTP and WTA estimates from each of the six groups will be tested for 'normality' using the Sparipo-Wilk method. Assessing the normality of the data is important when it comes to deciding whether to use parametric or non-parametric hypothesis tests to compare the different groups.

4.1.1 The Data Sets: Descriptive Statistics

As detailed in chapter 3, there were six main treatment groups involved in the experiment, each of which completed a slightly different variation of a CV survey which aimed to assess how much value participants place on the attribute of *certified organic cotton*.

Valuation data was elicited from two of groups of participants under the specification that the group's average stated value would determine the provision of the attribute (see section 3.3.1). These two *binding* treatment groups (WTP_B and WTA_B) were also asked to predict the average valuation elicited from the other group (i.e. WTP_B was asked to estimate the average value stated by the WTA_B group, and vice versa).

Two treatment groups provided valuation data under a purely *hypothetical* CV context, identical to the binding treatment except that no monetary consequence to participants' stated values was specified (WTP_H and WTA_H). These groups were also asked to estimate the average valuation of the two binding groups.

The final two treatment groups were not asked to state their own valuation for the good, but were asked only to predict the average valuation estimate elicited from the two binding groups (WTP_P and WTA_P).

The raw WTP and WTA data sets collected from the binding and hypothetical treatment groups are presented in Appendix 12 and the six prediction data sets can be found in Appendix 13. Summary statistics derived from these data sets are presented below.

Table 6. Hypothetical and Binding WTP/WTA - Descriptive Statistics

	WTP_B	WTA_B	WTP_H	WTA_H
Mean ($)	4.741	12.833	5.500	10.793
S.E. Mean	0.848	1.256	0.671	1.325
Median ($)	4	13.5	5	8
Std. Dev.	4.408	6.879	3.674	7.133
n	27	30	30	29

On average, participants in the un-endowed binding treatment group were willing to pay a maximum of $4.70 to have their non-organic cotton t-shirt upgraded to an identical organic cotton t-shirt (see table 6). The average minimum amount that the endowed binding group (WTA_B) demanded in compensation to exchange their organic cotton t-shirt for an identical non-organic cotton t-shirt

was $12.80. Since average WTP and WTA both exceeded the per-capita cost of the organic cotton attribute ($C = $4.20), both groups received the organic cotton t-shirts as specified by the provisional rule (see section 3.3.4), and the WTP participants were each required to pay the $4.20 before they would receive their t-shirt.

The median WTP and WTA for the two binding groups are $4 and $13.50 respectively. Median measures of central tendency are sometimes used in CV analysis because they are less influenced by outliers in the data set, which is often an issue in such investigation.

The mean WTP and WTA of the hypothetical treatment groups are $5.50 and $10.80 respectively. From these value measures alone it appears that a hypothetical bias may be present in this data, with participants overstating what they would be willing to pay to gain organic cotton and understating how much they would demand to part with it in a hypothetical context. However, without the appropriate hypothesis tests this theory cannot yet be confirmed.

The median WTP and WTA of the hypothetical groups are $5 and $8 respectively, and appear to more closely converge than the equivalent median measures from the binding treatments ($4 and $13.50 respectively). This could indicate that hypothetical CV experiments produce smaller WTA-WTP gaps than those that include a binding provisional rule. Again, this observation cannot be validated or dismissed without the appropriate hypothesis tests.

Table 7. Predicted WTP/WTA - Descriptive Statistics

	WTP_P1	WTP_P2	WTP_P3	WTA_P1	WTA _P2	WTA_P3
Mean	6.125	2.690	3.233	12.419	8.267	4.630
S.E. Mean	0.781	0.823	0.747	1.207	1.206	0.945
Median	5	1	2	10	6.5	1
Std. Dev.	4.420	4.433	4.091	6.722	6.607	4.908
n	32	29	30	31	30	27

Notes:

WTP_P1: The WTP_P group's prediction of WTP_B

WTP_P2: The WTA_H group's prediction of WTP_B

WTP_P3: The WTA_B group's prediction of WTP_B

WTA_P1: The WTA_P group's prediction of WTA_B
WTA_P2: The WTP_H group's prediction of WTA_B
WTA_P3: The WTP_B group's prediction of WTA_B

Predicted WTP and WTA values varied considerable across the different treatment groups (see table 7). The two main prediction groups, whose sole task was to estimate the average WTP and WTA of the binding groups appears to have produced reasonably accurate estimations, with a mean estimated-WTP of $6.13 (WTP_P1) and a mean estimated-WTA of $12.42 (WTA_P1). These estimated values both seem similar to the actual average WTP and WTA of the binding groups. This may offer support to the idea that a group of un-endowed participants are able to produce unbiased estimates of the actual WTP and WTA data elicited from binding treatment groups.

The prediction estimates produced by the two hypothetical groups appear slightly less accurate however, with a mean estimated-WTP of $2.70 (WTP_P2) and a mean estimated-WTA of $8.30 (WTA_P2). Similarly, the two binding groups also appear far less able to estimate each other's WTP and WTA when compared to the two primary prediction groups, producing a mean estimated-WTP of $3.23 (WTP_P3) and a mean estimated-WTA of $4.63 (WTA_P3).

As previously noted, although the descriptive statistics presented in tables 6 and 7 may appear to support of oppose the various hypotheses detailed in section 3.2.3, statistical hypothesis testing is required before any judgements can be made.

4.1.2 Demographic Characteristics of the Treatment Groups

This section will look at the demographic make-up of the six treatment groups, comparing their sample statistics to the average sample statistics of the experiment group as a whole. The purpose of this analysis is to ensure that none of the treatment groups differ significantly in terms of their participant's characteristics, as such a difference may result in response biases between treatments which would be problematic.

Table 8. Demographic Characteristics by Treatment Group

Treatment	Total	WTP_P	WTA_P	WTP_H	WTA_H	WTP_B	WTA_B
Average Age	25.576	24.422	26.548	24.638	27.155	24.796	25.883
Male (%)	0.511	0.531	0.452	0.483	0.483	0.593	0.533
Student (%)	0.757	0.806	0.710	0.700	0.759	0.846	0.733
[a]Income	93.645	88.281	95.565	99.583	94.397	92.593	91.667
[b]Dependents	0.417	0.452	0.519	0.261	0.583	0.348	0.321
[c]Education	0.500	0.500	0.516	0.414	0.621	0.556	0.400
[b]Ethnicity	0.545	0.531	0.548	0.621	0.414	0.407	0.733
n	179	32	31	30	29	27	30

[a] Average weekly disposable income
[b] Average number of dependents
[c] Percentage with a degree or higher
[d] Percentage of European

Table 8 presents the demographic data collected from the 179 participants involved in the experiment. The *Total* column provides the average sample statistics for the experiment group as a whole, while the following six columns separate the sample statistics by treatment group. This allows for the identification of any divergence between the sample statistics of the experiment group as a whole and each of the individual treatment groups. Any such divergence would indicate that the treatment groups were not all equal in terms of participant characteristics, which could cause issues when it comes to comparing the data collected from each group.

As shown in the *Total* column, the average age of participants was just under 26, the proportion of males to females was roughly 50/50, 75% of participants were students, the average number of dependents per participant was less than 1, 50% of respondents had at least a bachelors degree, and just over 50% were European.

At first glance, it appears that the sample statistics from each of the six treatment groups more or less converge with that of the sample averages for the entire experiment, shown in the *Total* column. For example, the average age of respondents in the six treatments ranges from 24.42 in the WTP_P group to 27.15

in the WTA_H group, which are both relatively close to the sample average of 25.58.

However, to provide conclusive evidence that the participant groups do not differ significantly from the sample average, each of the treatment group's seven sample statistics are subtracted from the corresponding statistic from the *Total* column, and the t-statistic for this difference is calculated. For example, the average age of respondents in the WTP_P group (24.42) is subtracted from the average age of respondents in the entire experiment sample (25.58). This difference is 1.16, and with a t-statistic of 0.687, the null hypothesis that the difference between the two means is zero is not rejected at the 5% level of significance, implying that the WTP_P does not differ significantly from the sample average in terms of average age. The complete table of results are presented Appendix 14.

The only treatment group which differs in any significant way from the sample average is the WTA_B group who had 19% less European participants than the experiment sample as a whole. While this difference was significant at the 5% level, the discrepancy is not enough to imply that the data collected from this group will be biased, as the six other demographic characteristics are all in line with the sample averages.

The seven demographic characteristics of the five other treatment groups were not different statistically from the sample average in any significant way (Appendix 14). It can therefore be concluded that all six of the treatment groups are identical in terms of the characteristics of their participants, and any differences in valuation results between the groups cannot be attributed to biases resulting from difference in the sample.

4.1.3 Testing for Reliability and Theoretical Validity

According to Mitchell and Carson, "reliability refers to the extent to which the variance of the WTP [or WTA] amounts given by respondents in a contingent valuation survey is due to random sources, or 'noise'" (1989, p. 211). Reliability can be assessed by regressing the WTP or WTA values elicited from participant against the demographic characteristics of those participants to ensure that the variance in value estimates is not purely random. It is suggested that if a

regression model, using only a few key variable, produces an R^2 greater than 0.15, then the reliability of the data is confirmed.

Regressing the *binding* WTP and WTA values on various combinations of independent variables (demographic characteristics of participants) using the ordinary least squares method produced the following linear models (standard errors in parentheses):

$$\underset{}{\text{WTP\_B} = 3.187 + 3.813(\text{Ethnicity}^{1})}$$
$$(1.013) \quad (1.587)$$

$$\text{WTA\_B} = 19.936 - 6.554(\text{Male}^{2}) - 6.531(\text{Student}^{3}) + 1.612(\text{Ethnicity})$$
$$(3.256) \quad (2.074) \qquad (2.370) \qquad (2.370)$$

Note: Standard errors of coefficients are in the parentheses

[1]Ethnicity: binary variable where 1 = European and 0 = non-European

[2]Male: binary variable where 1 = male and 0 = female

[3]Student: binary variable where 1 = Bachelors degree or higher and 0 = less than a Bachelors

The WTP_B model presented above produced an adjusted R^2 of 0.155, which implies that variation in the independent variable (ethnicity) can account 15.5% of the variation in the dependent variable (WTP). This is a relatively low R^2, but it is still larger than the 0.15 which Mitchell and Carson suggest is necessary to establish reliability.

The WTA_B model presented above produced an adjusted R^2 value of 0.342, implying that variation in the independent variables can account for around 34% of the variation in the dependent WTA variable. This is well above Mitchell and Carson's 0.15 lower R^2 limit, and confirms the reliability of the binding WTA data.

The reliability of the binding WTP and WTA data is confirmed, with the adjusted R^2 value of each regression model exceeding 0.15. This indicates that the independent variables included in each of the regression models are able to

explain enough of the variation in WTP or WTA to establish that this variation is not entirely due to random 'noise'.

4.1.4 Calculating the Consumer Benefit of Organic Cotton

To demonstrate what a researcher may use real-world CV data for, the binding WTP data for the *organic cotton* attribute will now be used to reveal the demand curve, and to calculate the level of 'benefit' that provision of the organic cotton attribute would bring to the sample group.

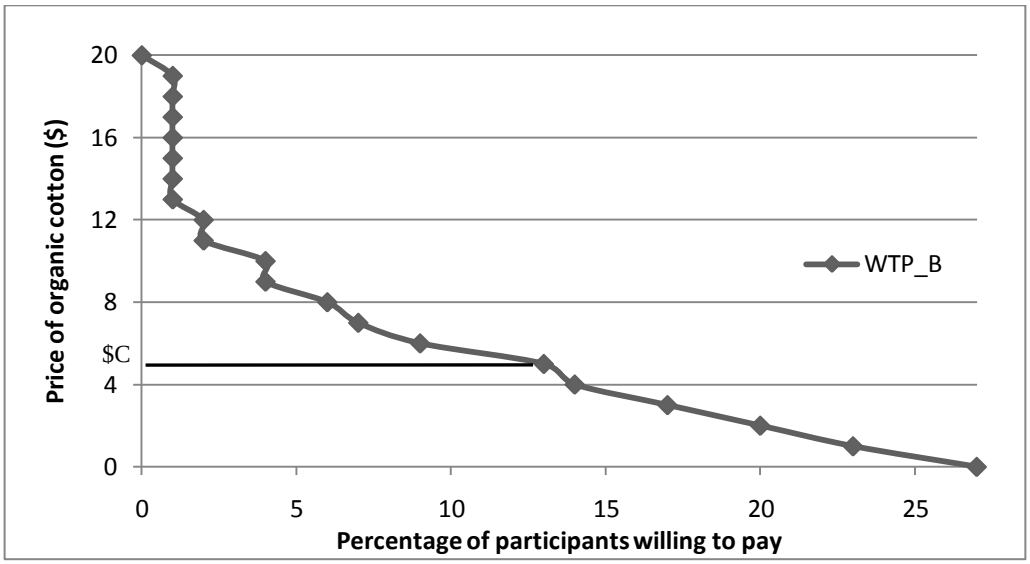


Figure 6. Demand for Organic Cotton

Figure 6 shows the inverse relationship between the bid price for organic cotton and the number of participants in the binding treatment who were willing to pay that amount to gain the attribute. We can see that at a price level of $20 none of the participants were willing to pay to gain the good. As the price level decreases the number of participants willing to pay to gain the good increases, as we would expect. The actual price of the organic cotton is shown by the line $C and is $4.20. This value for organic cotton was the difference in price between the plain non-organic cotton t-shirts and the plain organic cotton t-shirts that were to be given to participants in the binding WTP and WTA treatments. Using this price level and the demand curve of the binding WTP group, it is possible to estimate the benefit to the participants of receiving the organic cotton attribute.

The "traditional measure of consumer benefit, proposed by Dupuit in the nineteenth century and championed by Marshall, is consumer surplus, which is defined as the area under the ordinary (Marshallian) demand curve and above the price line (Bateman et al., 2002, p. 23). The consumer surplus is a measure of the difference between what people are *willing to pay* to gain a good, and what they are *required to pay* for that good.

For the WTP_B group, this consumer surplus can be calculated by subtracting the price of organic cotton ($4.2) from the maximum WTP of each participant whose WTP is greater than $4.2, and summing these results. As shown in Appendix 15, the total consumer surplus for the WTP_B group is $51.40. This indicates that if the organic cotton attribute was provided only to those participants who were willing pay at least $4.20 to gain it, the total consumer benefit of this provision would equal just under $51.40.

This consumer benefit measure relies on the assumption that those participants not willing to pay $4.20 for organic cotton do not have to do so. However, since the contingent valuation scenario described in section 3.2.1 defined the organic cotton attribute as a public good, specifying that it would either be provided to all participants for the price $C or withheld from all participants, the true measure of the good's benefit to the group is reduced. To calculate the true benefit to the group or collectively purchasing the organic cotton attribute, the benefit measure needs to account for the 'losses' experiences by those participants who are forced to pay an amount greater than their maximum WTP.

To calculate the true consumer benefit of the organic cotton attribute, the cost $C ($4.20) is subtracted from the average WTP of all the respondents ($4.74) which gives us the average consumer surplus of $0.54. This average consumer surplus is then multiplied by the total number of respondents in the treatment group (n=27) to give the total consumer surplus measure, which equals $14.58. While the additional benefit that the binding WTP group would receive from organic cotton is lessened when taking into account those participants not willing to pay for the good, there is still a positive consumer surplus associated with the good.

The average consumer surplus for organic cotton for our sample group, calculated previously, would be multiplied by the total number of people affected by the proposed provision of the good to estimate the total benefit to society that the good would create. This is possible only if the sample group is randomly selected from the defined population, and the significance of the estimated average WTP allows for the estimation of population parameters.

4.1.4.1 Generalising the Consumer Surplus Measure

In a real world application of the CV method, it is likely that the researcher would want to extend the consumer surplus measure derived from their participant group to a larger population in order to make inferences about the total economic gain associated with a proposed change. To support this generalisation, the researcher would need to ensure that the participant group used to elicit WTP data is a random representative sample of the defined population, and is of a large enough size to produce a statistically significant average WTP measure.

If these conditions are met, then the researcher would typically multiple the average consumer surplus calculated for their sample group by the total number of people in the population who would be affected, estimating the total consumer surplus of the proposed change. However, since the sample group used for the research was not randomly selected from some larger defined population, such inferences cannot be made using any of the WTP of WTA data collected.

4.1.5 Testing Data for Normality

In the following sections of analysis and discussion, the 10 data sets which were collected from the six treatment groups (Appendix 12 and 13) will be compared to one another to test the hypotheses outlined in section 3.2.3.These hypothesis tests will be carried out using the parametric technique of comparing sample means and analysing the associated t-statistic. However, this method of hypothesis testing requires that the population distribution for each data set follows the normal distribution curve (Stock & Watson, 2007). While it is generally assumed that most sample distribution will approximate to a normal curve if the number of observations is large enough, we will test each of our six

data sets using the Sharipo-Wilk test to determine whether they can in fact be considered ‘normal’.

This technique was chosen to assess the normality of the data due the relatively small sample sizes of each data set. The “Shapiro-Wilk Test is more appropriate for small sample sizes (< 50 samples)”, while an alternative test called the Kolmogorov-Smirnov test is generally used for data sets with over 2000 observations (Leard Statistics, 2010).

Table 9. Shapiro-Wilk Test for Normality

	Shapiro-Wilk		
	W-Statistic	**df**	**Sig.**
WTP_B	.877	27	.004
WTA_B	.895	30	.006
WTP_H	.940	30	.088
WTA_H	.869	29	.002
WTP_P1	.815	32	.000
WTP_P2	.656	29	.000
WTP_P3	.782	30	.000
WTA_P1	.899	31	.007
WTA _P2	.896	30	.007
WTA_P3	.746	27	.000

Note: The Shapiro-Wilk test was conducted using SPSS statistical software

Table 9 presents the results of the Shapiro-Wilk test conducted using SPSS for all 10 of the data sets collected from the 6 treatment groups. The WTP_H data set is the only set which follows a normal distribution curve according to the Shapiro-Wilk test, with a significance value greater than 0.05. All nine of the other data sets have a significance level of less than 0.05 which means that the null hypothesis that the data is normally distributed is rejected at the 5% level of significance.

Since a majority of the data sets do not conform to the normal distribution pattern, standard parametric hypothesis tests are not ideal for use with this data. Therefore, two non-parametric hypothesis test – Mood’s median test and the Mann-Whitney U test – will be used alongside the parametric t-test when

comparing the data sets. Further detail regarding each of these hypothesis tests will be provided in the following analysis of the binding WTA and WTP data sets.

4.2 **Identifying the WTA-WTP Gap**

The following analysis will compare the WTA and WTP data elicited from the two binding treatment groups (WTA_B and WTP_B) in order to answer the first research question detailed in section 3.2.3: *Does a significant WTA-WTP disparity emerge in a CV experiment even when controlling for features of weak experimental design?*

4.2.1 Parametric and Non-Parametric Hypothesis Tests

Since the normality of the two data sets cannot be confirmed (Shapiro-Wilk test, $p < 0.05$) both parametric and nonparametric hypothesis tests will be used to analyse the data. The t-statistic for the difference between two population means is the first hypothesis-testing method which is used, as well as the nonparametric Mann-Whitney U test and Mood's Median test. Each of these methods aims to identify whether the two individual data sets could have theoretically been collected from an identical population distribution or whether there is a significant difference between the likely population distributions of the two sets. The process for calculating the relevant test statistic for each of the three hypothesis testing methods will be detailed below using the two binding data sets (summarised in table 10 below). For the remainder of the analysis sections however, SPSS software will be used to estimate the test statistics.

Table 10. Sample Statistics of the Binding WTA and WTP Groups

Value Measure	**Mean**	**Median**	**n**
WTA_B	12.833	13.5	30
	(6.879)		
WTP_B	4.741	4	27
	(4.408)		
Difference	8.092	9.5	

Note: Sample standard deviations in parenthesis

4.2.1.1 Parametric t-Test

The first method used to test whether there is a significant difference between the WTP_B and WTA_B groups involves using the mean, standard deviation, and number of observations of each data set to estimate the population distribution of each set and determine whether the distributions are equal.

The null hypothesis that we are interested in, that WTA_B = WTP_B, can be reinterpreted as H_0: $\mu_{WTA\_B} - \mu_{WTP\_B} = d_0$, where μ_{WTA_B} is the population mean of the binding WTA group, μ_{WTP_B} is the population mean of the binding WTP group, and d_0 = is the difference between the two means (Stock & Watson, 2007, p. 82). Since the true population mean of each data set is unknown, the sample mean for each group is used as an estimator: $\text{Y-bar}_{WTA\_B} - \text{Y-bar}_{WTP\_B} = d_0$. The statistical significance of d_0 is determined by the *t-statistic* for the difference between the two means, which relies on the standard error (SE) of the estimator for the difference between the population means: $\text{Y-bar}_{WTA\_B} - \text{Y-bar}_{WTP\_B}$

$$\text{SE}(\text{Y-bar}_m - \text{Y-bar}_w) = \sqrt{\left(\frac{s.d_m^2}{n_m} + \frac{s.d_w^2}{n_w}\right)} \qquad (4)$$

$$\text{t-statistic} = \frac{(\text{Ybar}_m - \text{Ybar}_w) - d_0}{\text{SE}\ (\text{Ybar}_m - \text{Ybar}_w)} \qquad (5)$$

Table 10 shows that the difference between the two sample means is equal to \$8.10 indicating that the average WTA stated by the binding group was \$8.10 higher than the average WTP stated by the binding group. The standard error of this difference can be calculated using the standard deviations and number of observations of each data set:

$$\text{SE}(\text{Y-bar}_{WTA\_B} - \text{Y-bar}_{WTP\_B}) = \sqrt{\left(\frac{6.879^2}{30} + \frac{4.408^2}{27}\right)}$$

S.E = 1.515

$$\text{t-statistic} = \frac{8.10 - 0}{1.515}$$

t = 5.340

According to the student-t distribution for 30+ degrees of freedom (Stock & Watson, 2007, p. 757), the t-statistic calculated above (5.340) exceeds the 1-sided critical value of 2.75 which allows us to reject the null hypothesis that the two data sets are equal at a 1% level of significance. Furthermore, the t-statistic is positive, so we can infer that the mean of WTA_B is significantly greater than the mean of WTP_B.

As previously noted, this method of hypothesis testing assumes that the population distribution for each sample is normally distributed, which our previous analysis was unable to confirm. The following two hypothesis testing techniques make no such assumption about the population distribution.

4.2.1.2 Mann-Whitney U Test

The Mann-Whitney U test "may be used to test whether two independent groups have been drawn from the same population... [and] is one of the most powerful of the nonparametric tests" (Siegel, 1956, p. 116). The null hypothesis, H_0, is that that the two populations of WTA_B and WTP_B have the same distribution. The alternate hypothesis, H_1, is that WTA_B is stochastically larger than WTP_B. The alternate hypothesis can be accepted if the probability that a score from WTA_B (wta_b) is larger than a score from WTP_B (wtp_b) is greater than one-half: p(wta_b > wtp_b) > 0.5 (Siegel, 1956, p. 116). H_1 could be also be p(wta_b > wtp_b) < 0.5, to test if WTA_B is less than WTP_B, or p(wta_b > wtp_b) ≠ 0.5 for a two-tailed test.

The first stage in calculating the U-statistic is to pool the WTA_B and WTP_B data sets, and rank each data point from the lowest to the highest (e.g. 1, 2, 3...), before separating the values back into their two original groups and summing the assigned ranks for each group. See Appendix 16 for this process. The sum of assigned ranks for the WTA group is 1137.5 (n_1=30), while the sum of assigned ranks for the WTP group is 515.5 (n_2=27).

The U statistic is then calculated using the following formula (Siegel, 1956, p. 120):

$$\boldsymbol{U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1} \qquad \textbf{(6)}$$

$$U = 30 \times 27 + \frac{30(30+1)}{2} - 1137.5$$

$$U = 137.5$$

Since the smallest of the sample groups (WTP_B) still has more than 20 data points, the significance of the U-statistic can be calculated by substituting the U value (137.5) into the following formula to produce a z-score (Siegel, 1956, p. 123):

$$z = \frac{\mathrm{U} - \frac{n_1 n_2}{2}}{\sqrt{\frac{(n_1)(n_2)(n_1+n_2+1)}{12}}} \qquad (7)$$

$$z = \frac{127.5 - \frac{30 \times 27}{2}}{\sqrt{\frac{(30)(27)(30+27+1)}{12}}}$$

$z =$ -4.275

According to the normal distribution table, the z-value of -4.275 is outside the range of -4.0 to 4.0, which means that it has a one-tailed probability of occurring if H_0 is true of p <0.0003 (Siegel, 1956, p. 247). Therefore, the null hypothesis that the population distribution of WTA_B and WTP_B are equal can be rejected at a 1% level of significance, and we can confirm that stated WTA is stochastically larger than stated WTP in the two binding treatments.

4.2.1.3 Mood's Median Test

Mood's median test "is a procedure for testing whether two independent groups differ in central tendencies. More precisely, the median test will give information as to whether it is likely two independent groups... have been drawn from populations with the same median" (Siegel, 1956, p. 111). Because this hypothesis test is based on the median of the sample distribution rather than the mean, it is less influenced by samples with large outliers. The null hypothesis is that the WTA_B and WTA_B data sets are both drawn from populations with the same median, while the alternate hypothesis is that the WTA_B data is drawn

from a population with a median larger than that from which the WTP_B data was drawn (single-tailed test).

The first step in this hypothesis test is to pool the two sets of data and calculate the resulting median, which for the WTA_B and WTP_B groups is equal to 8. Next, the number of data points in each sample group that are less than or equal to this combined median (8), are counted, as are the number of observations greater than the combined median. These results are presented in table 11 below.

Table 11. Median Test: Form for Data

	WTP_B	WTA_B	Total
Scores > combined median (8)	4 (A)	22 (B)	26
Scores ≤ than combined median (8)	23 (C)	8 (D)	31
Total	27	30	N = 57

If the two samples are from populations with the same median, then we would expect around half of each group's scores to be above the combined median, and half to be below the combined median. To test whether there is a significant difference between the allocation of scores between the four segments of table 11, we use the Chi-Squared test, assuming the combined sample groups have more than 40 observations (Siegel, 1956, p. 112). Using the formula presented below (8), the χ^2 value for the data can be calculated.

$$\chi 2 = \frac{N(|AD-BC|-\frac{N}{2})^2}{(A+B)(C+D)(A+C)(B+D)} \qquad (8)$$

Note: A, B, C, and D correspond to each of the four elements of table 11, and N is the combined sample size.

$$\chi 2 = \frac{57(|4 \times 8 - 22 \times 23| - \frac{57}{2})^2}{(4+22)(23+8)(4+23)(22+8)}$$

$$\chi 2 = 17.328$$

The standard Chi Square table of critical values (Siegel, 1956, p. 249) shows that an $\chi^2 \geq 10.83$ with 1-degree of freedom, has the probability of occurring if H_0 is true of $p<0.0005$ for a 1-tailed test. H_0 is therefore rejected at a 1% level of significance, and it can conclude that the median of the WTA_B population is significantly greater than the median of the WTP_B population.

These three hypothesis tests (parametric t-test, the Mann-Whitney U test, and Mood's median test) will all be used to test each of the hypotheses outlined in section 3.23. However, for the remainder of the hypothesis tests, the statistical software SPSS will be used to conduct the calculations and the results will be summarised as shown in table 12 in the following discussion section.

4.2.2 The WTA-WTP Gap: Results and Discussion

Table 12. Comparison of WTA and WTP from the Binding Treatment Groups

H_0: WTA_B = WTP_B

H_1: WTA_B > WTP_B

Null: there is no significant difference between the value estimates collected from the WTA_B and WTP_B treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTA_B	12.833	13.5	30
	(6.879)		
WTP_B	4.741	4	27
	(4.408)		
Difference	8.092	9.5	
[a]**t-stat**	5.340**		
[b]**U**	137.5**		
[c]χ^2		17.328**	

Notes: Sample standard deviations in parenthesis

[a]Parametric t-test

[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (1-tailed test)
**H_0 rejected at 1% level of significance (1-tailed test)

The null hypothesis that the binding data collected from the WTA_B and WTP_B groups could have come from identical population distributions is rejected at the 1% level of significance according to all three of the hypothesis tests provided above. This indicates that population distribution from which the WTA_B data was collected has a significantly greater mean and median that the population distribution of the WTP_B data.

Accordingly, those participants endowed with organic cotton at the beginning of the experiment (WTA_B) value the attribute nearly 3 times more highly than those participants not endowed with the good (WTP_B). This WTA/WTP ratio of 2.7, is strikingly similar to that observed by Bischoff who calculated a ratio, under similar experimental condition, of 2.9 for a publically available tutorial session (2008, p. 292).

4.2.2.1 Weak Experimental Design

From the results presented above we can conclude that that a significant WTA-WTP gap exists within the data collected from the two binding treatment groups. It is also apparent that this gap is statistically significant even though these CV surveys both used an incentive compatible elicitation mechanism and provisional rule, removed potential hypothetical bias by including a binding monetary outcome, and ensured participant anonymity. These results indicate that the WTA-WTP disparity observed in our data is not likely the result of weak experimental design, but rather a fundamental difference in the value that each group places on the organic cotton attribute.

As noted in chapter 3, the experimental design adopted for the thesis was based on that used by Bischoff (2008), and included the same initial allocation of property rights, the same binding provisional rule, an identical elicitation mechanism which even used the same question structure, and involved a good which was also defined as being a *public good*. It is therefore interesting that while actual binding valuation estimates differ considerable between this current research and Bischoff's research, the WTA/WTP ratio in both experiments are

nearly identical (2.7 in this current experiment and 2.9 in Bischoff's experiment). This finding is a little troubling, because if two experiments which adopt identical survey designs produce the same WTA/WTP ratio even though they are valuing different goods using different participant samples, then the observed disparity could theoretically be attributed to some unaccounted for flaw in the methodology.

4.2.2.2 Mean versus Median WTA/WTP Ratios

It may be expected that the median measure of WTA and WTP could result in a lower disparity since this measure of central tendency is less vulnerable than the sample mean to large outliers in the data which are common in CV experiments, especially when using a open-ended elicitation format. Horowitz and McConnell's meta analysis provided support for this theory, finding that of the "41 experiments that reported ratios of both means and medians... the ratio involving means was greater than the ratio involving medians in close to 80% of the experiments" (2002, p. 430). However, contrary to this finding, the WTA/WTP ratio from our data is actually greater when calculated using median measures (3.3) rather than mean measures (2.9).

4.2.2.3 WTA/WTP Ratio and the Substitutability Theory

The organic cotton attribute was defined in the CV experiment as being a *public good* where consumption was non-rival and no participant could be excluded from purchasing the good (or selling it) once the final level of provision was decided. A common theme in the CV literature, which was summarised by Horowitz and McConnell (2002), is that WTA/WTP ratios are typically much higher for public goods than they are for ordinary private goods. The average ratio for a public good in Horowitz and McConnell's study was 10.41, considerably larger than the average ratio for an ordinary private good, 2.92 (2002, p. 433).

This theme is typically explained in the literature as resulting from the substitutability effect (see section 2.7) which suggests that a good which has easily accessible substitutes will have lower WTA/WTP ratio than a good which has fewer available substitutes, due to differences in the diminishing marginal rate of substitution between these types of goods. Accordingly, the effects of unanticipated ownership utility and parting disutility on the WTA-WTP gap

should be different for public and private goods. This idea is one of the reasons why Bischoff's experiment (2008) and this current experiment both defined the good to be valued as a *public good*, to assess the contribution of unanticipated ownership utility and parting disutility to the WTA-WTP gap of a the less-substitutable good.

However, the results from this experiment, as well as Bischoff's experiment, appears to suggest that participants may actually have viewed the good being valued as an ordinary private good rather than a public good. This judgement is made since the WTA/WTP ratio in both Bischoff's experiment and this current experiment (2.9 and 2.7 respectively) appear to resemble the typical ratio for an ordinary private good (2.92) rather the ratio for a public good (10.41), as specified by Horowitz and McConnell (2002). Therefore, assessing the relative contributions of unanticipated ownership utility and parting disutility to the endowment effect for a public good, is not possible.

4.3 Unanticipated Ownership Utility and Parting Disutility

This analysis is related to the research question: *Are individuals who are not endowed with a good able to anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good?*

This section follows the framework constructed by Loewenstein and Adler (1995) and revised by Bischoff (2008), which aims to isolate the two main effects that these authors identify as contributing to the endowment effect: unanticipated ownership utility and parting disutility (loss aversion). As previously noted, the technique used to separate these two effects involves collecting actual WTA and WTP values as well as predicted WTA and WTP values, and then comparing the sample distributions of the actual and predicted WTA data to assess whether the two data sets are significantly different.

If predicted WTA is equal to actual WTA, this indicates that the un-endowed prediction group is able to fully anticipate the ownership utility gained from possessing the good, and therefore the endowment effect (i.e. the difference between actual WTA and WTP) is entirely due to parting disutility and other conventional factors. Conversely, if predicted WTA was to is equal actual WTP, this would suggest that participants in the prediction group were unable to anticipate any of the ownership utility that they would gain from possessing the good. The main assumption underlying this analysis is that the predicted WTP is able to provide an unbiased estimate of actual WTP. To test this assumption, the WTP_B and WTP_P1 data sets are compared, and the difference between distribution statistics are tested in table 13 below.

4.3.1 Testing Predicted WTP for Bias

Table 13. Comparison of WTP Binding and WTP Prediction Treatment Groups

H_0: WTP_B = WTP_P1
H_1: WTP_B ≠ WTP_P1

Null: there is no significant difference between the value estimates collected from the WTP_B and WTP_P1 treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTP_B	4.741	4	27
	(4.408)		
WTP_P1	6.125	5	32
	(4.42)		
Difference	-1.384	-1	
[a]**t**	-1.200		
[b]**U**	343.000		
[c]χ^2		0.022	

Notes: Sample standard deviations in parenthesis
[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

The null hypothesis that the binding WTP data and the predicted WTP data could have been drawn from the same population distribution cannot be rejected by at 5% level of significance according to the three test results provided in table 13. The two-tailed parametric t-statistic (t = -1.2), suggests that while average predicted-WTP ($6.13) was slightly greater than the actual average WTP, this difference is not large enough to be considered statistically significant at a 5% level ($p > 0.05$). The non-parametric Mann-Whitney U statistic (U = 343) also confirms that H_0 cannot be rejected at any reasonable level of significance. Furthermore, Mood's median test produces an χ^2 value of 0.187 which fails to be statistically significant at either the 1% or 5% level of confidence. We can conclude from these three hypothesis tests, that the true population distributions of predicted and actual WTP do not differ significantly in either mean or median value measures.

This indicates that the predicted WTP data and the actual WTP data are not statistically different, providing strong support for the core assumption underlying the following analysis – that participants are able to provide an unbiased estimate of true WTP.

4.3.2 Testing Predicted WTA for Bias

Table 14. Comparison of WTA Binding and WTA Prediction Treatment Groups

H_0: WTA_B = WTA_P1
H_1: WTA_B ≠ WTA_P1

Null: there is no significant difference between the value estimates collected from the WTA_B and WTA_P1 treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTA_B	12.833	13.5	30
	(6.879)		
WTA_P1	12.419	10	31
	(6.722)		
Difference	0.414	3.5	
[a]**t**	0.238		
[b]**U**	442.500		
[c]χ^2		0.146	

Notes: Sample standard deviations in parenthesis
[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

The null hypothesis that the binding WTA data and the predicted WTA data could be drawn from an identical population distribution cannot be rejected by at 5% level of significance according to any of the three test statistics provided in table 14. The two-tailed parametric t-statistic (t = 0.238), the Mann-Whitney U statistic (U = 442.5), and Mood's Chi Square statistic (χ^2 = 0.407) are all considered statistically insignificant at the 5% level. This means that the distribution of the WTA_P1 and WTA_B data sets are not considered significantly different, suggesting that the participants in the un-endowed prediction group were able to fully anticipate the endowment effect that was

experienced by those participants who were endowed with the organic cotton attribute (WTA_B).

This finding is contrary to what Loewenstein and Adler (1995) and Bischoff (2008) found in their respective experiments. These authors both determined that estimated WTA was significantly lower that actual WTA, attributing this disparity to the *unanticipated ownership utility* effect by suggesting that the participants in the estimation group were unable to anticipate the full utility gained by those participants in the actual WTA group. These authors then used a simple mathematical equation to determine the proportion (β) of the WTA-WTP gap (in the actual elicitation groups) that could be attributed to the unanticipated ownership effect.

Although the finding of this current research do not suggest that any unanticipated ownership utility effect is present in the binding WTA group (since WTA_B = WTA_P1), we will now calculate the beta value for this effect, just as Bischoff (2008) had done so.

4.3.3 Calculating the Effect of Unanticipated Ownership Utility

To calculate the proportion of the WTA-WTP gap that can be attributed to unanticipated ownership utility, signified by β, the following formula is used:

$$\beta = \frac{(WTA - WTA\_P)}{(WTA - WTP)} \qquad (1)$$

Table 15. Mean WTA and WTP from the Binding and Prediction Groups

Group	N	Mean	S.E of Mean
WTA_B	30	12.833	1.256
WTP_B	27	4.741	0.848
WTA_P1	31	12.419	1.207
WTP_P1	32	6.125	0.781

Substituting the values presented in table 15, into the formula (1), we calculate a β value of 0.051.

$$\beta = \frac{(12.833 - 12.419)}{(12.833 - 4.741)}$$

$$\beta = 0.051$$

This β of 0.051 suggests that the unanticipated ownership utility effect only accounts for around 5% of the total WTP-WTA disparity, leaving other conventional factors, such as the parting disutility effect, to account for the remaining 95%. This β is substantially lower than that observed by Loewenstein and Adler who calculated a β value 0.94 (1995, p. 935), and also much lower than that calculated by Bischoff, β = 0.24 (2008, p. 293).

4.3.4 Discussion

The reason for the insignificant level of unanticipated ownership utility observed in this current research is unclear. It could be argued that the significant biases in predicted WTA observed by the two previous research experiments (Loewenstein and Adler, 1995; Bischoff, 2008) was due to factors such as differences in the demographic characteristics of the treatment groups, rather than the unanticipated ownership utility. This may indicate that the unanticipated ownership utility effect does not actually exist, and that the past findings related to this theory are inaccurate.

On the other hand, it is possible that this current research did not allow the ownership utility effect to develop fully in the binding WTA group. Considering that these participants (WTA_B) had not actually been given their organic cotton t-shirt at the time they stated their WTA, it is possible that they did not actually feel they 'owned' the good. If participants had physical possession of the t-shirt, like the respondents in Loewenstein and Adler (1995) held possession of their mugs, then the unanticipated ownership utility effect may have been more significant for organic cotton, resulting in an even larger WTA-WTP gap.

A third explanation for why the β value in this current experiment is insignificant compared to that observed in previous studies, is that the organic

cotton attribute may simply not provide individuals with any unanticipated ownership utility. Since this attribute is intangible, both the WTP and WTA groups are essentially endowed with the same good (a t-shit), so the fact that the WTA group's t-shirts were made from organic cotton may not have significantly increased the utility felt by these participants. This would suggest that the WTA-WTP gap observed in this experiment is most likely caused by the parting disutility effect rather than an unanticipated ownership utility effect.

4.4 Hypothetical Bias and the WTA-WTP Gap

The section of analysis will compare the WTA and WTP data elicited from the two binding treatment groups (WTA_B and WTP_B) to the corresponding data sets obtained from the hypothetical treatment groups (WTA_H and WTP_H) to answer the following two research questions:

a) *Does a purely hypothetical CV experiment produce significantly different valuation results to those obtained from an identical experiment but where participants' decisions have a binding monetary consequence?*

b) *Does a purely hypothetical CV experiment produce a significantly different WTA-WTP gap than an identical experiment with a binding monetary consequence?*

4.4.1 Identifying a Hypothetical Bias

Table 16. Comparison of WTA Binding and WTA Hypothetical Treatment Groups

H_0: WTA_B = WTA_H
H_1: WTA_B ≠ WTA_H

Null: there is no significant difference between the value estimates collected from the WTA_B and WTA_H treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTA_B	12.833 (6.879)	13.5	30
WTA_H	10.793 (7.133)	8	29
Difference	2.04	5.5	
[a]**t**	1.118		
[b]**U**	356.500		
[c]χ^2		0.857	

Notes: Sample standard deviations in parenthesis
[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

From the three hypothesis tests presented in table 16 (above), it is apparent that the data collected from the hypothetical WTA group does not differ in any significant way from the data collected from binding WTA data. Therefore, the null hypothesis cannot be rejected at any reasonable level of significance (5%).

Although the binding group's average WTA of $12.83 is slightly greater than the hypothetical group's average WTA, $10.79, according to the parametric t-test and the Mann-Whitney U test, this difference of $2.04 is not statistically significant at the 5% level. Furthermore, the low χ^2 value for Mood's median test indicates that the two data sets could have been collected from two populations with identical medians, so the binding and hypothetical data cannot be considered different in this regard either.

We can therefore conclude that introducing a binding monetary provisional rule, which connects participant's decisions to a real monetary payment, does not significantly affect how participants responded to the WTA survey in our experiment.

Table 17. Comparison of WTP Binding and WTP Hypothetical Treatment Groups

H_0: WTP_B = WTP_H
H_1: WTP_B ≠ WTP_H

Null: there is no significant difference between the value estimates collected from the WTP_B and WTP_H treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTP_B	4.741	4	27
	(4.408)		
WTP_H	5.5	5	30
	(3.674)		
Difference	-0.759	-1	
[a]**t**	-.702		
[b]**U**	339.000		
	135		

[c]χ^2	0.569

Notes: Sample standard deviations in parenthesis
[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

The hypothesis tests presented in table 17 indicate that the null hypothesis - that the hypothetical and binding WTP data sets collected from our sample groups are equal - cannot be rejected at the 5% level of significance. While average hypothetical WTP ($5.50) is slightly larger than the average binding WTP ($4.74), all three of the parametric and nonparametric hypothesis tests that were use conducted confirm that this difference is insignificant.

We can therefore conclude that, just as in the WTA comparison (table 16 introducing a binding monetary provisional rule does not significantly affect how participants responded to the WTP questions presented in our organic cotton CV survey.

4.4.2 Hypothetical Bias and the WTA-WTP Gap

Table 18. Comparison of WTA Hypothetical and WTP Hypothetical Treatment Groups

H_0: WTA_H = WTP_H
H_1: WTA_H > WTP_H

Null: there is no significant difference between the value estimates collected from the WTA_H and WTP_H treatment groups.

Value Measure	**Mean**	**Median**	**n**
WTA_H	10.793 (7.133)	8	29
WTP_H	5.5 (3.674)	5	30
Difference	5.293	3	
[a]**t**	3.565**		

[b]**U**	246.500**	
[c]χ^2		4.886*

Notes: Sample standard deviations in parenthesis
[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
*H_0 rejected at a 5% level of significance (1-tailed test)
**H_0 rejected at 1% level of significance (1-tailed test)

The results presented in table 18 (above) indicate that the hypothetical WTA data is significantly larger than the hypothetical WTP data, both in terms of mean and median measures. The null hypothesis that these two data sets could have been drawn from identical population distributions is rejected at the 1% level of significance by the t-test (t = 3.565) and Mann-Whitney U test (U = 246.5), and is rejected at the 5% level of significance by Mood's median test (χ^2 = 4.886). It can therefore be confirmed that a statistically significant WTA-WTP disparity exists within the hypothetical CV data.

The value of the WTA-WTP gap, in terms of sample averages, observed in the hypothetical data is $5.29 which is slightly less than the gap observed in the binding data: $8.10. Furthermore, the WTA/WTP ratio is slightly smaller in the hypothetical groups than the binding groups (1.96 compared to 2.7 respectively). However, since no hypothetical bias was observed in our organic cotton CV experiment, either in terms of a WTP or WTA, then it can be suggested that purely hypothetical CV experiments may not produce significantly different WTA-WTP gaps than experiments involving real monetary transactions. However, this finding may be a unique feature of this CV design, and further research would be needed before this result could be generalised to all CV experiments.

4.5 **Prediction Bias across Treatments**

This section of analysis aims to compare the actual WTP and WTA values from the binding treatment groups to the predicted WTP and WTA values elicited from all six of the different treatment groups. This analysis aims to assess how the prior experience with the CV scenario, which was experienced by the hypothetical and binding treatment groups, impacts on the ability of these participants to produce unbiased valuation estimates.

To review the process through which the six different prediction data sets were collected see section 3.3.3. Essentially the different prediction data sets can be summarised as follows:

1) **WTP_P1** – participants are only asked to predict WTP_B's average response (i.e. WTP_P treatment)
2) **WTP_P2** – Hypothetical WTP data is elicited from participants before they are asked to predict WTP_B's average response (i.e. WTA_H treatment)
3) **WTP_P3** – Binding WTP data is elicited from participants before they are asked to predict WTP_B's average response (i.e. WTA_B treatment)

The process for the three WTA prediction groups is the same as above but P2 and P3 groups were collected from WTP_H and WTP_H, respectively.

In all of the hypotheses tests presented so far in this analysis section, significance levels to reject the null hypothesis have been limited to 5% and 1%. This was because included lesser levels of significance, such as a 10% rejection region, would not have changed the results of any of the hypothesis tests (i.e. p-values for each test was either very low or very high). However, the hypothesis tests outlined in this section include a 10% rejection measure, as well at the typical 5% and 1% levels. This is because some of the null hypotheses cannot be rejected at the 5% level, but they can be rejected at the 10% level, so including this higher measure is relevant.

4.5.1 WTP Prediction Comparisons

Table 18. Comparison of WTP Binding and the Three WTP Prediction Groups

H_0: WTP_B = WTP_P1/P2/P3
H_1: WTP_B ≠ WTP_P1/P2/P3

Null: there is no significant difference between the value estimates collected from the WTP_B and WTP_P1/P2/P3 treatment groups.

	WTP_P1	WTP_P2	WTP_P3
Actual Mean (WTP_B)	4.741	4.741	4.741
Predicted Mean	6.125	2.69	3.233
Actual Median	4	4	4
Predicted Median	5	1	2
Difference in Means	-1.384	2.051	1.508
Difference in Medians	-1	3	2
t-statistic	-1.200	1.735 '	1.334
U	343.000	240*	296.5’
χ^2	0.022	3.473 ‘	0.398
n	32	29	30

[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
‘H_0 rejected at a 10% level of significance (2-tailed test)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

From table 18 we can see that, as examined in section 4.3.1, the first prediction group (WTP_P1) whose only task was to estimate the average WTP stated by the binding group, is able to provide an unbiased estimate of actual WTP. None of the three hypothesis tests were able to reject the null at any meaningful level of significance.

However, for the WTP_P2 data set, elicited from the WTA_H treatment group, the null is rejected at the 10% level of significance by both the t-test (t = 1.735) and Mood’s median test (χ^2 = 3.473), and is rejected at the 5% level by the Mann-Whitney U test (U = 240). This indicates that the predicted WTP stated by

the hypothetical WTA group ($2.69) cannot be considered an unbiased estimate for actual WTP ($4.74).

The third set of predicted WTP data (WTP_P3), collected from the WTA_B group, cannot be considered an unbiased estimate for actual WTP. The null hypothesis that WTP_B is equal to WTP_P3 is rejected at a 10% level of significance by the Mann-Whitney U test (U = 295.5), though the other two hypothesis tests are unable to reject the null. Regardless of the insignificant t-statistic and χ^2 value, the null hypothesis is still rejected by the nonparametric U test, confirming that the two data distributions are significantly different from one another.

From the comparison of prediction estimate presented above, it appears that participants are able to provide an unbiased estimate of the average WTP of the binding treatment group as long as that they are not first asked to state their own WTA valuation for organic cotton (in either the hypothetical or binding treatment). Some aspect present in both the hypothetical and binding WTA elicitation process appears to introduce a significant bias into the predicted WTP stated by these participants.

The WTP prediction bias appears to be downwards in direction, with the WTP_P2 and WTP_P3 groups both providing an average estimate lower than the actual WTP mean. However, the direction of bias is not confirmed by this analysis.

4.5.2 WTA Prediction Comparison

Table 19. Comparison of WTA Binding and the Three WTA Prediction Groups

H_0: WTA_B = WTA_P1/P2/P3
H_1: WTA_B ≠ WTA_P1/P2/P3

Null: there is no significant difference between the value estimates collected from the WTA_B and WTA_P1/P2/P3 treatment groups.

	WTA_P1	WTA_P2	WTA_P3
Actual Mean (WTA_B)	12.833	12.833	12.833
Predicted Mean	12.419	8.267	4.63

Actual Median	13.5	13.5	13.5
Predicted Median	10	6.5	1
Difference in Means	0.414	4.566	8.203
Difference in Medians	3.5	7	12.5
t-statistic	0.238	2.623‘	5.221**
U	442.5	281.5’	114**
χ^2	0.146	2.469	12.879**
n	31	30	27

[a]Parametric t-test
[b]Mann-Whitney U test
[c]Mood's median test (χ^2 corrected for continuity)
‘H_0 rejected at a 10% level of significance (2-tailed test)
*H_0 rejected at a 5% level of significance (2-tailed test)
**H_0 rejected at 1% level of significance (2-tailed test)

Similar to the previous WTP prediction comparison, table 19 shows that participants are able to produce an unbiased estimate for actual WTA under the condition that they are not first asked to state their own valuation for organic cotton. The null hypothesis that binding WTA is equal to predicted WTA is not rejected for the WTA_P1 group, but is rejected at the 10% level for the WTA_P2 group (t-stat = 2.623’ and U-stat = 281.5’), and at the 1% level for WTA_P3 (t-stat = 5.221**, U-stat = 114**, and χ^2 = 12.879**).

This finding is consistent with that of the analysis in section 4.3.2, and suggests that eliciting WTP valuation data from participants, whether through a hypothetical or binding CV survey, prior to the prediction exercise introduces a significant bias into these groups’ ability to predict the WTA of the binding group.

4.6 **Results Summary**

Section 4.1 of the previous analysis chapter confirmed that the six treatment groups from which the experimental data was collected do not differ significantly in terms of the demographic characteristics of respondents. Therefore, any differences in the valuation data collected from these groups can be directly attributed to the intentional crafted differences in each group's survey design.

It was calculated that the mean WTP and WTA of the binding treatment groups were \$4.74 (S.E = 0.848) and \$12.83 (S.E = 1.256) respectively. The reliability of the binding WTP and WTA responses was confirmed by regressing several explanatory variables derived from participants' demographic characteristics on each set of valuation data. This regression analysis confirmed that variation in binding WTP and WTA responses was not entirely due to random 'noise' but could be, in part, explained by variation in the characteristics of the participants.

The binding WTP data was used to construct a demand curve for organic cotton, calculating that the total consumer surplus that the group would gain if the organic cotton attribute was provided to them all at the price of \$4.20, would be \$14.58. This consumer surplus would increase \$51.40 if only those participants who were willing to pay more than \$4.20 had to purchase the organic cotton attribute.

Lastly, the Sharipo-Wilk test determined that all of the data sets collected from the six experimental groups, with the exception of the WTA_H data, could not be considered to follow a normal distribution at the 5% level of significance. It was decided to included nonparametric hypothesis tests, as well as the standard parametric t-test, in the data analysis process

Section 4.2 outlines the process of testing for a statistically significant difference between the distribution of two independent sets of data using the parametric t-test, the Mann-Whitney U test, and Mood's median test. These three hypothesis tests all confirmed that our binding WTA data set was significantly larger than the WTP data set at a 1% level of significance. This confirmed that a

significant WTA-WTP gap, with an absolute value of around $8 and a ratio of 2.7:1, was observed in the binding experimental groups even though these treatments controlled for elements of *weak experimental design* which are commonly used to explain such a disparity. This analysis confirmed that a significant WTA-WTP disparity does emerge in a CV experiment even when controlling for features of weak experimental design.

It was also noted that the ratio of binding WTA/WTP (2.7), while similar to the ratio identified in Bischoff's (2008) experiment (2.9), appears to more closely resemble the typical ratio for a ordinary private good (2.92) rather than a public good (10.41). It was therefore concluded that participants may have viewed the organic cotton attribute as a private good rather than as a public good, even though the provision of the attribute was designed to make it resemble a public good.

Section 4.3 compared the actual WTP and WTA values elicited from the binding groups to the estimated WTP and WTA collected from the two primary prediction groups, to identify the impact that the unanticipated ownership utility effect had on the WTA-WTP gap. The null hypothesis that the average predicted WTP ($6.13) and actual average WTP ($4.42) were equal was unable to be rejected at any meaningful level of significance, and predicted WTP was therefore confirmed as an unbiased estimator for actual WTP.

It was then determined that predicted WTA (mean = $12.42) and actual WTA (mean $12.83) were also not significantly different in terms of their value distributions. This implied that predicted WTA was an unbiased estimator for actual WTA, a finding which opposed Loewenstein and Adler (1995) and Bischoff (2008) who both determined that their prediction groups were unable to provide an unbiased estimate for actual WTA. These two authors explained this bias in WTA predictions as resulting from an unanticipated ownership utility effect which is experienced by the actual WTA group but not the predicted WTA group. However, the results of this current experiment suggest that the unanticipated ownership utility does is not present in the binding WTA valuation data and is unlikely to have contributed to the WTA-WTP gap.

It is therefore concluded that the significant WTA-WTP gap observed in our experiment is the result of parting disutility and other conventional factors, rather than an unanticipated ownership utility effect.

Section 4.4 compared the WTA and WTP data elicited through the two binding treatments to the WTA and WTP data elicited through the hypothetical treatments to assess whether a significant *hypothetical bias* is present, and whether this bias has any significant impact on the size of the WTA-WTP gap. It is shown that hypothetical WTA did not differ in any statistically significant way from binding WTA. This finding was also true for the two WTP data sets. It was therefore concluded that the hypothetical nature of a CV experiment is not likely to produce statistically different WTA and WTP results compared to those obtained from a binding experiment with real monetary outcomes.

Furthermore, while the two hypothetical value measures did produce a substantial WTA-WTP gap, this gap was not significantly different from the gap observed in the binding treatments. This finding was based on the fact that neither hypothetical WTA nor WTP differed from their binding equivalents.

Section 4.5 provides an analysis of the six different prediction data sets that were collected, comparing each set's distribution to that of the data it was attempting to estimate (i.e. WTP_B or WTA_B). It was concluded that the main prediction groups for both WTP and WTA (WTP_P1 and WTA_P1) were able to provide unbiased estimates of actual WTP and WTA. However, the remaining four prediction data sets, collected from the hypothetical and binding elicitation groups after they each completed their main valuation experiments, was found to provide significantly biased estimates of actual WTP and WTA.

The null hypothesis that the WTP_P2 and WTP_P3 data sets were identical to the WTP_B data set was rejected at the 5% level for the P2 group and at the 10% level for the P3 group. Furthermore, the null that the WTA_P2 and WTA_P3 data sets were identically distributed to the WTA_B data set was also rejected, at the 10% level for the P2 group and at the 1% level for the P3 group.

It was concluded from these findings that participants in our experiment were able to provide unbiased estimates of binding WTP and WTA, provided they

were not asked to state their own valuation of the good prior to the prediction exercise. If the participants were asked to state their WTP or WTA in either the hypothetical or binding treatment, before being asked to predict the actual valuation data, their ability to provide an unbiased estimate is eliminated.

Chapter 5 Conclusion

This aim of this thesis was to construct and conduct a simple CV experiment which was designed specifically to test for difference in valuation estimates that arise due to particular aspects of the survey's design. As noted in the introductory remarks, the main features of experimental design that this thesis was investigating were: 1) the initial allocation of property rights for the good being valued and the phrasing of the elicitation questions, 2) the hypothetical nature of a contingent valuation scenario. Within these two areas of investigation four main research questions were posed, each of which will now be concluded with reference to the findings of the organic cotton CV experiment outlined in chapter 3.

5.1 The Endowment Effect

Does a significant WTA-WTP disparity emerge in a CV experiment even when controlling for features of weak experimental design?

Based on the results of our experiment, we can conclude that a significant WTA-WTP disparity does emerge in a CV experiment, even when features of weak experimental design are controlled for. A statistically significant WTA-WTP gap is observed for organic cotton, with participants in the two binding CV treatments stating an average WTA to give up the attribute of $12.83 and an average WTP to gain the attribute of $4.75. Since the binding CV surveys were both specifically designed to eliminate other potential sources of the WTA-WTP gap that were identified in the literature, such as *weak experimental design*, the results of this thesis indicate that the disparity is most likely due to an endowment effect experienced by those participants given initial property rights to the good. This *endowment effect* suggests that participants who are endowed with a good (WTA) place a greater real value on that good than participants who are not endowed with the good (WTP), and the WTA-WTP gap is caused by this difference in real, rather than flaws in the survey methodology.

The implications of this finding is that when designing a CV experiment, the initial setting of property rights and the way in which the elicitation questions are phrased will have a significant impact on the value estimates produced. Therefore, researchers need to think carefully when designing their CV

experiment, to assess whether the population effected by the proposed change being investigated currently hold property rights for the good, or do not hold property rights for the good.

Are individuals who are not endowed with a good able to anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good?

Our results confirm that individuals who are not endowed with a good (predicted WTA) are able to fully anticipate the magnitude of the endowment effect experienced by participants who are endowed with the good (binding WTA), provided that this un-endowed group are not asked to state their own valuation for the good prior to the prediction process. This conclusion is derived from the findings detailed in section 4.3 and section 4.5, which showed that predicted WTA is considered an unbiased estimator for actual WTA, and that predicted WTP also matches actual WTP.

This conclusion also provides evidence that the endowment effect is most likely caused by parting disutility rather than unanticipated ownership utility. Furthermore, this finding provides strong evidence to suggest that participants are able to provide an unbiased estimate for the preferences of others', adding to the relatively small amount of literature on this topic.

5.2 **Hypothetical Bias**

Does a purely hypothetical CV experiment produce significantly different valuation results to those obtained from an identical experiment where participants' decisions have a binding monetary consequence?

This thesis provides compelling evidence against the idea that the hypothetical nature of a CV experiment has a significant influence on valuation results. As shown in section 4.4, the value estimates obtained through the binding and hypothetical treatment groups do not differ in any significant way. This means that participants in a hypothetical CV experiment should respond to elicitation questions in a similar way to how they would respond if the experiment involved real monetary transactions. Therefore, criticism of the CVM based

around the notion that the hypothetical nature of this valuation technique makes it invalid, appears to be unfounded.

Does a purely hypothetical CV experiment produce a significantly different WTA-WTP gap than an identical experiment with a binding monetary consequence?

Lastly, the results of this experiment fail to show that the hypothetical nature of a CV study has any impact on the magnitude of the WTA-WTP gap. This means that while a significant disparity between these two measures of value is observed in hypothetical CV experiments, the gap cannot be attributed in any way to the fact that participant's decision do not have binding monetary consequences. This conclusion provides additional support for the endowment effect explanation of the WTA-WTP gap, as appose to the weak experimental design argument.

References

Adamowicz, W., Dupont, D., Krupnick, A., & Zhang, J. (2011). Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. *Journal of Environmental Economics and Management, 61*(2), 213-226.

Adamowicz, W. L., Bhardwaj, V., & Macnab, B. (1993). Experiments on the Difference between Willingness to Pay and Willingness to Accept. *Land Economics, 69*(4), 416-427.

Anderson, D. M. (2010). Estimating the economic value of ice climbing in Hyalite Canyon: An application of travel cost count data models that account for excess zeros. *Journal of Environmental Management, 91*(4), 1012-1020.

Bateman, I., J., Carson, R., T., Day, B., Hanemann, W., M., Hanley, N., Hett, T., et al. (2002). *Economic valuation with stated preference techniques: A manual*: Cheltenham, U.K. and Northampton, Mass.:

Elgar; distributed by American International Distribution Corporation, Williston, Vt.

Bestard, A. B., & Font, A. R. (2009). Environmental diversity in recreational choice modelling. *Ecological Economics, 68*(11), 2743-2750.

Bischoff, I. (2008). Endowment effect theory, prediction bias and publicly provided goods: an experimental study. *Environmental and Resource Economics, 39*(3), 283-296.

Bischoff, I., & Meckl, J. (2008). Endowment effect theory, public goods and welfare. *Journal of Socio-Economics, 37*(5), 1768-1774.

Bishop, R., C., & Romano, D. (1998). *Environmental resource valuation: Applications of the contingent valuation method in Italy*: Kluwer Academic Publishers.

Brown, G., Jr., & Hammack, J. (1973). Dynamic Economic Management of Migratory Waterfowl. *Review of Economics and Statistics, 55*(1), 73-82.

Brown, T. C., Peterson, G. L., & Tonn, B. E. (1995). The values jury to aid natural resource decisions. *Land Economics, 71*(2), 250.

Bryan, S., & Jowett, S. (2009). Hypothetical versus real preferences: results from an opportunistic field experiment. *Health Economics, 19*(12), 1502-1509.

Camacho, C., Eva, García, G., Aurora, Georgantzís, N., & Sabater, G., Gerardo. (2004). An Experimental Validation of Hypothetical WTP for a Recyclable Product. *Environmental and Resource Economics, 27*(3), 313-335.

Carson, R. C. M. a. R. T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the Dictator. *American Economic Review, 92*(4), 1218-1221.

Coursey, D. L., Hovis, J. L., & Schulze, W. D. (1987). The Disparity between Willingness to Accept and Willingness to Pay Measures of Value. *Quarterly Journal of Economics, 102*(3), 679-690.

Cummings, R. G., Harrison, G. W., & Rutstrom, E. E. (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible? *American Economic Review, 85*(1), 260-266.

Driml, S. (2002). Travel Cost Analysis of Recreation Value in the Wet Tropics World Heritage Area. *Economic Analysis and Policy, 32*(2), 11-26.

Du Preez, M., Tessendorf, S., & Hosking, S. G. (2010). Application of the Contingent Valuation Method to Estimate the Willingness-to-Pay for Restoring Indigenous Vegetation in Underberg, Kwazulu-Natal, South Africa. *South African Journal of Economic and Management Sciences, N.S., 13*(2), 135-157.

Fremling, G. M., & Posner, R. A. (1999). Market Signaling of Personal Characteristics. University of Chicago.

Gregory, R., & Furby, L. (1987). Auctions, Experiments and Contingent Valuation. *Public Choice, 55*(3), 273-289.

Hanemann, W. M. (1991). Willingness to Pay and Willingness to Accept: How Much Can They Differ? *American Economic Review, 81*(3), 635-647.

Hanemann, W. M. (1994). Valuing the Environment Through Contingent Valuation. *The Journal of Economic Perspectives, 8*(4), 19-43.

Hoehn, J. P., & Randall, A. (1987). A satisfactory benefit cost indicator from contingent valuation. *Journal of Environmental Economics and Management, 14*(3), 226-247.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior, 7*(3), 346-380.

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

Jones-Lee, M. W., Hammerton, M., & Philips, P. R. (1985). THE VALUE OF SAFETY: RESULTS OF A NATIONAL SAMPLE SURVEY. *Economic Journal, 95*(377), 49-72.

Kahn, A., & Tice, T. E. (1973). Returning a favor and retaliating harm: The effects of stated intentions and actual behavior. *Journal of Experimental Social Psychology*, *9*(1), 43-56.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, *98*(6), 1325-1348.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives, 5*(1), 193-206.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.

Kealy, M. J., Montgomery, M., & Dovidio, J. F. (1990). Reliability and predictive validity of contingent values: Does the nature of the good matter? *Journal of Environmental Economics and Management, 19*(3), 244-263.

Knetsch, J. L. (1990). Environmental policy implications of disparities between willingness to pay and compensation demanded measures of values. *Journal of Environmental Economics and Management, 18*(3), 227-237.

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics, 99*(3), 507-521.

Lipton, D. W., Wellman, K., Sheifer, I. C., Weiher, R. F., & NOAA, N. C. O. P. O. (1995). *Economic valuation of natural resources [microform] : a handbook for coastal resource policymakers*.

Silver Spring, Md. :: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Coastal Ocean Office.

List, J. A., & Gallet, C. A. (2001). What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environmental and Resource Economics, 20*(3), 241-254.

Loehman, E., & De, V. H. (1982). Application of Stochastic CHoice Modeling to Policy Analysis of Public Goods: A case Study of Air Quality Improvements. *Review of Economics & Statistics, 64*(3), 474.

Loewenstein, G., & Adler, D. (1995). A Bias in the Prediction of Tastes. *Economic Journal, 105*(431), 929-937.

Loomis, J. B. (1990). Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. *Journal of Environmental Economics and Management, 18*(1), 78-85.

Mankiw, N. G. (2007). *Principles of Economics [Fourth Edition]* (Fourth Edition ed.): Thompson Higher Education

Manson, OH

Mansfield, C. (1999). Despairing Over Disparities: Explaining the Difference Between Willingness to Pay and Willingness to Accept. *Environmental and Resource Economics, 13*(2), 219-234.

Merino, A. (2003). Eliciting Consumers Preferences Using Stated Preference Discrete Choice Models: Contingent Ranking versus Choice Experiment.

Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method*: Washington, D.C.: Resources for the Future.

Morrison, G. C. (2000). WTP and WTA in repeated trial experiments: Learning or leading? *Journal of Economic Psychology, 21*(1), 57-72.

Onwujekwe, O., & Nwagbo, D. (2002). Investigating starting-point bias: a survey of willingness to pay for insecticide-treated nets. *Social Science & Medicine, 55*(12), 2121-2130.

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization, 65*(3-4), 703-713.

Plott, C. R., & Zeiler, K. (2005). The Willingness to Pay-Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review, 95*(3), 530-545.

Randall, A., & Stoll, J. R. (1980). Consumer's Surplus in Commodity Space. *American Economic Review, 70*(3), 449-455.

Sayman, S., & Öncüler, A. (2005). Effects of study design characteristics on the WTA-WTP disparity: A meta analytical framework. *Journal of Economic Psychology, 26*(2), 289-312.

Schmidt, U., & Traub, S. (2009). An Experimental Investigation of the Disparity between WTA and WTP for Lotteries. *Theory and Decision, 66*(3), 229-262.

Schuman, H., & Presse, S. (1981). Questions and Answers in Attitude Survey. *New York: Academic Press*.

Shaw, W. D. (2005). The Road Less Traveled: Revealed Preference and Using the Travel Cost Model to Value Environmental Changes. *Choices, 20*(3), 183-188.

Shogren, J. F., Shin, S. Y., Hayes, D. J., & Kliebenstein, J. B. (1994). Resolving Differences in Willingness to Pay and Willingness to Accept. *American Economic Review, 84*(1), 255-270.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York:: McGraw-Hill.

Soliño, M., Prada, A., & Vázquez, M. X. (2010). Designing a forest-energy policy to reduce forest fires in Galicia (Spain): A contingent valuation application. *Journal of Forest Economics, 16*(3), 217-233.

Stock, J. H., & Watson, M. W. (2007). *Introduction to Econometrics [Second Edition]*: Addison Wesley, Boston.

Taylor, Morrison, M., & Boyle, K. (2010). Exchange Rules and the Incentive Compatibility of Choice Experiments. *Environmental and Resource Economics, 47*(2), 197-220.

Taylor, L. O. (1998). Incentive Compatible Referenda and the Valuation of Environmental Goods. *Agricultural and Resource Economics Review, 27*(2), 132-139.

Van Boven, L., Loewenstein, G., & Dunning, D. (2003). Mispredicting the endowment effect:: Underestimation of owners' selling prices by buyer's agents. *Journal of Economic Behavior & Organization, 51*(3), 351-365.

Veisten, K. (2007). Contingent valuation controversies: Philosophic debates about economic theory. *Journal of Socio-Economics, 36*(2), 204-232.

Venkatachalam, L. (2004). The contingent valuation method: a review. *Environmental Impact Assessment Review, 24*(1), 89-124.

Wardman, M. (1988). A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour. *Journal of Transport Economics and Policy, 22*(1), 71-91.

Whitehead, J. C., Pattanayak, S. K., Van Houtven, G. L., & Gelso, B. R. (2008). Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science. *Journal of Economic Surveys, 22*(5), 872-908.

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

Appendix 1: Information on the Benefits of Organic Cotton

Introduction

We are interested in your opinion about t-shirts made using certified organic cotton. The T-Shirt that you have been given is made using non-organic cotton. However, this shirt can be made using 100% certified organic cotton, at an additional cost to you, depending on the outcome of this decision making experiment.

The Benefits of Organic Cotton

Cotton is traditionally known as the world's 'dirtiest' crop because of the large amount of synthetic fertilizers and insecticides used in its production. Cotton covers around "2.5% of the world's cultivated land yet uses 16% of the world's insecticides" (Organic Trade Association, 2009). These chemicals used to produce cotton can be very damaging to the native vegetation and animals where the cotton is grown, and their effect on the soil often means that no other plants will be able to survive in the area for many years after contamination.

Certified organic cotton is grown using methods and materials that have a low impact on the environment. Non-genetically modified cotton seeds, natural fertilizers, and crop rotation strategies are used to create an organic cotton production system which:

- Helps to replenish and maintain soil fertility,
- Reduces the use of toxic and persistent pesticides and fertilizers
- Builds biologically diverse agriculture

(Guerena & Sullivan, 2003)

<< Back | Next >>

Appendix 2: Instructions Given to the WTP_B Group

qualtrics.com

Instructions

A number of your fellow participants have also just been given a free non-organic cotton t-shirt, and a set of instructions identical to yours.

You will each be given a series of options where you will indicate whether you would prefer to keep the non-organic cotton t-shirt or exchange it for an identical certified organic cotton t-shirt for which you would be required to pay the amount indicated. For example, if the option says:

Please select your most preferred option from the two alternatives below

- ○ Keep non-organic cotton t-shirt, or
- ◉ Exchange for an identical **organic cotton** t-shirt and pay **$5**

and you choose the second option, you are indicating that you would be willing to pay $5.00 to have your shirt made from organic cotton.

T-shirts made using certified organic cotton cost more than t-shirts made using non-organic cotton. This additional cost of organic cotton, represented by **$C**, will be told to you once all experiment sessions have been completed.

Once all participants in this treatment group have completed their survey we will use the results to calculate the maximum amount of money that each of you are willing to pay to have your shirt made from certified organic cotton rather than non-organic cotton. From these values we will calculate the group's average willingness to pay for the organic option.

If the **average** amount that the group is willing to pay is **less** than $C, then all of your t-shirts will be made using non-organic cotton and no payment will be required, regardless of whether or not this is your personal preference.

However, if the **average** amount the group is willing to pay is **equal to or greater than** $C, then all of your t-shirts will be made using organic cotton and you will each be required to pay this additional cost ($C) before you are able to pick up your t-shirt.

Please remember that the results of this experiment are **binding**, and you will all receive either:

- the default option of having your t-shirt made from non-organic cotton, or
- the alternative option of an identical t-shirt made from 100%-certified organic cotton for which you will be required to pay the amount $C.

The option that you will receive is determined by the group as a whole and the choices you all make.

<< Back | Next >>

Appendix 3: Instructions Given WTA_B Group

qualtrics.com

Instructions

A number of your fellow participants have also just been given a free certified organic cotton t-shirt, and a set of instructions identical to yours.

You will each be given a series of options where you will indicate whether you would prefer to keep the certified organic cotton t-shirt or exchange it for an identical non-organic cotton t-shirt for which you would receive a personal payment of the amount indicated. For example, if the option says:

Please select your most preferred option from the two alternatives below

- ○ Keep organic cotton t-shirt, or
- ◉ Exchange for an identical non-organic cotton t-shirt and $5

and you choose the second option, you are indicating that you would be willing to exchange your certified organic cotton t-shirt for an identical non-organic cotton t-shirt and a personal payment of $5.00.

T-shirts made using certified organic cotton cost more than t-shirts made using non-organic cotton. This additional cost of organic cotton, represented by **$C**, will be told to you once all experiment sessions have been completed.

Once all participants within this treatment group have completed their survey we will use the results to calculate the **minimum** amount of money that each of you demand before you are willing to choose the non-organic t-shirt over the certified organic t-shirt. From these values we will calculate the a**verage minimum payment** demanded by the treatment group as a whole to choose the non-organic t-shirt option. This is called the 'willingness to accept' of the group.

If the **average** minimum payment demanded by the group is **less** than the $C, then all of your shirts will be made from non-organic cotton and you will each receive a payment of $C, regardless of whether or not this is your personal preference.

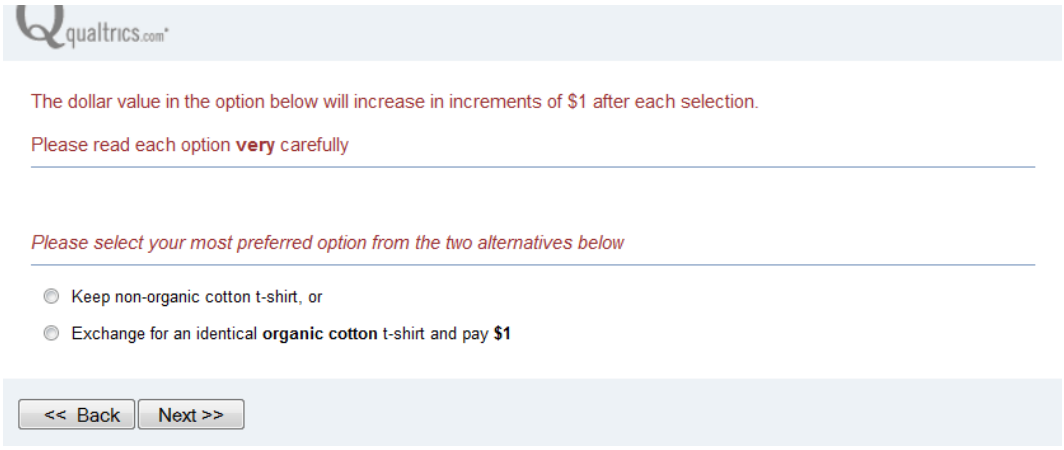
However, if the **average** minimum payment demanded by the group is **equal to or greater than** $C, then all of your t-shirts will be made using certified organic cotton, as originally offered, and no further payments will be made.

Please remember that the results of this experiment are **binding**, and you will all receive either:
- the default option of a free 100% certified organic cotton t-shirt, or
- the alternative option of an identical t-shirt made from non-organic cotton and a payment of $C.

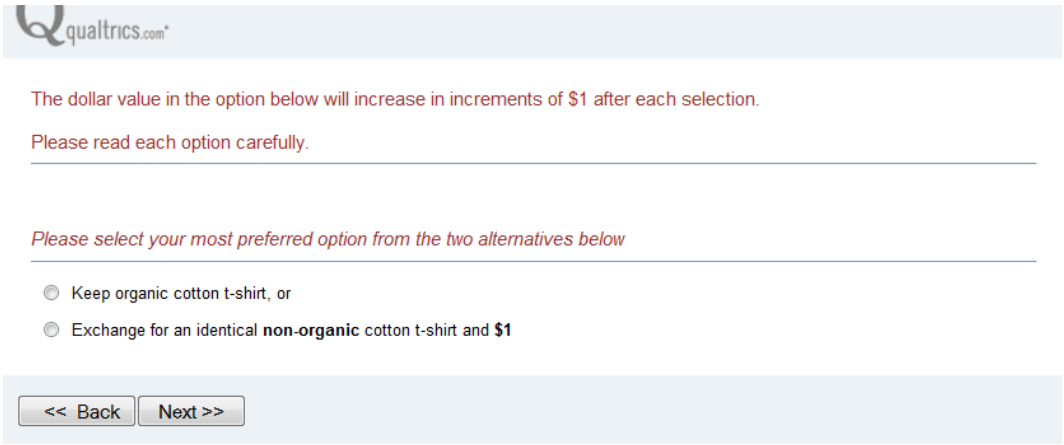
The option that you will receive is determined by the group as a whole and the choices you all make.

<< Back | Next >>

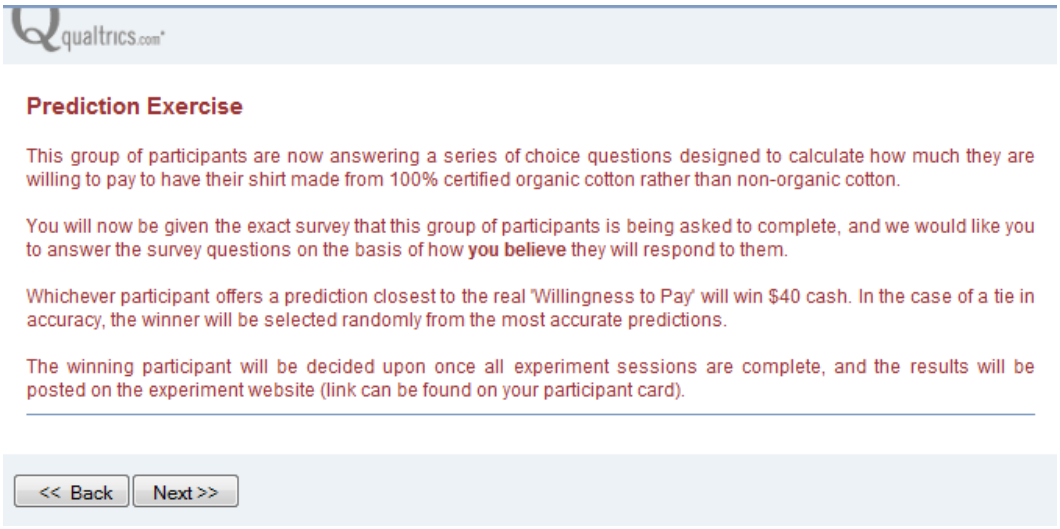
Appendix 4: First Elicitation Question for the WTP Group



Appendix 5: First Elicitation Question for the WTA Group



Appendix 6: Additional Information for WTP_P Group



Appendix 7: Information Sheet – Announcement Stage

Information at announcement stage

Waikato Management School
Te Raupoa

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

- **We are recruiting participants for an experiment.**
- **It will last about one hour and will take place in a computer lab.**
- **The experiment is about how people rank objects on the basis of their preferences and how they value different ways of producing cotton.**
- **To pay for your time if you take part you will be given $25 at the end of the experiment.**
-
- **To guarantee yourself a spot please e-mail to:**
-
- fdp1@waikato.ac.nz
-
- **And type "experiment" in the subject**

THANKS!!!!

Appendix 8: Participant Card

Participant Card

resp_id: 101
tr_wid: 1

http://www.mngt.waikato.ac.nz/ric/exp_results.html

Contact: fdp1@waikato.ac.nz

Security code: 10151

Appendix 9: Information Sheet for Participants

Information Sheet for Participants

Waikato Management School
Te Raupapa

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Preference Elicitation in a Private and Public Good Experiment
Department of Economics

Staff associated with the research:
Prof. Riccardo Scarpa rscarpa@waikato.ac.nz
Prof. Tony Marley ajmarley@uvic.ca
Dr. Graeme Doole gdoole@waikato.ac.nz
Mr. Francis Powley fdp1@waikato.ac.nz

Overview and purpose

Upon your completion of the experiment you will be given $25 cash. Please hold on to the card with your personal identifiers as it contains a web-link to a page where the results of this survey will be displayed as soon as possible. This experiment is about different types of cotton t-shirts and the way they are produced, particularly organically produced t-shirts. We are not only interested in the value that you place on these t-shirts, but also in the way you rank different types of t-shirts. Some of you will be given a t-shirt and the opportunity to have it changed depending on your decisions.

What the research will involve

The experiment is divided into two parts and will take roughly 30-45 minutes (maximum one hour) to complete. For the first part you will each be asked to repeatedly rank some t-shirts based on your personal preference towards each of the available options. The second part will have different forms. For some it will involve choosing the most preferred of two options in a series of repeated choice statements, again based around the product of a t-shirt. For the remainder of participants it will involve making predictions about the value that other students place on a particular attribute. To give you anonymity the experiment will be completed through a computerised questionnaire. Some personal information will be collected (e.g. sex, disposable income, etc.), with your permission, but such information will only be associated with your respondent ID number ("resp_id" in your card) and so will ANY response you give to the computerised survey.

The study itself will serve as the basis for a Thesis in Economics and potentially for an article to inform other researchers on the results obtained. The final Thesis will be accessible through the Waikato University Library to all students and staff, and any journal articles will be viewable by those subscribed to the journal.

Declaration to participants

If you take part in the study, you have the right to:

- Refuse to answer any particular question, and to withdraw from the study before the end of the lab session. However, in this case you will forgo the payment which is only received for survey completion.
- Ask any further questions about the study that occurs to you during your participation.
- Be given access to a summary of the findings from the study when it is concluded.

If you have any questions about this research project you can contact Francis Powley on 027 254 3484 or e-mail to fdp1@students.waikato.ac.nz.
This project is being supervised by Riccardo Scarpa and he can be contacted at (07) 838 4848 or RSCARPA@mngt.waikato.ac.nz

I agree with the terms above ☐

Please enter the ID number you were assigned

Start the survey

Appendix 10: Introductory Dialogue

Hello and welcome. Thank you all for coming.

My name is Francis Powley, and this is Professor Ric Scarpa, and we will be coordinating this experiment session today.

Before we begin, it is important that you have all read through and understood the **Participant Information Sheet** that should be on the screen in front of you.

Has everyone read through that?

.....

Once you have read through the Information Sheet, and if you are happy with the conditions it lays out, tick the 'Agree Box' and enter the ID Number from your participant card, labelled resp_id.

This participant number will be the only thing connecting you to the answers you provide. This insures total anonymity.

Please read all of the information in the experiment, paying particular attention to the instructions you are given, and think about your answer carefully before making a selection. If you think you have made a mistake you can always go back and correct your answer. Once you have completed the first set of choice tasks a link will appear which will lead you on to the second part of the experiment.

Once you reach the second part you will need to enter you Participant ID number again, and also your treatment code, labelled tr_wid.

Again, please read all of the information you are given thoroughly and think about each question before you answer. Once you complete the second half of the experiment you will be given your final instructions and you will be able to collect your $25.

All further instructions regarding your choice tasks are included in the online questionnaires. If you have any technical issues or questions regarding the tasks please raise your hand and either myself or Ric will come over to help.

Remember, we are interested in your personal preference towards each of the options you are given in this choice experiment. So there is no right or wrong answers.

Appendix 11: Waikato Management School Ethics Approval

Research Office
Waikato Management School
The University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand

Amanda Sircombe
Research Manager

Phone +64 7 838 4376
Fax +64 7 838 4063
Email amandas@waikato.ac.nz
www.management.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

MANAGEMENT SCHOOL
Te Raupapa

MEMO

To	Professor Riccardo Scarpa, Department of Economics
From	Amanda Sircombe, Research Manager
Date	6th December 2010
Subject:	**Waikato Management School Ethical Application**

Dear Ric

Ethics Application WMS 10/214
Differences in ranking outcomes under alternative forms of guidance and preferences for organic cotton textiles

The above project, as outlined in this ethics application, has been granted Ethical Approval for Research by the Waikato Management School Ethics Committee.

Please note: should you make changes to the project as outlined in the approved ethics application, you may need to reapply for ethics approval.

Best wishes for your research

Regards

Amanda Sircombe
Research Manager

Appendix 12: Binding and Hypothetical WTP and WTA – Raw Data

WTP_B	WTA_B	WTP_H	WTA_H
0	1	0	1
0	1	0	1
0	1	0	2
0	1	0	3
1	6	0	4
1	7	2	5
1	7	3	5
2	8	4	5
2	10	4	6
2	10	4	6
3	10	4	6
3	10	5	7
3	10	5	8
4	10	5	8
5	12	5	8
5	15	5	10
5	15	6	10
5	15	6	10
6	15	6	11
6	16	6	15
7	18	6	16
8	20	8	20
8	20	8	20
10	21	8	21
10	21	10	21
12	21	10	21
19	21	10	21
	21	10	21
	21	10	21
	21	15	
n = 27	n = 30	n = 30	n = 29

Notes:
WTP_B = WTP data collected from the binding treatment group
WTA_B = WTA data collected from the binding treatment group
WTP_H = WTP data collected from the hypothetical treatment group
WTA_H = WTA data collected from the hypothetical treatment group

Appendix 13: Predicted WTP and WTA – Raw Data

WTP_P	WTP_P2	WTP_P3	WTA_P1	WTA _P2	WTA_P3
0	0	0	1	1	1
2	0	0	1	1	1
2	0	0	3	1	1
3	0	0	4	1	1
3	0	0	5	1	1
3	0	0	5	1	1
3	0	0	7	1	1
3	0	0	7	1	1
4	0	0	8	3	1
4	0	0	9	3	1
4	0	0	10	5	1
4	0	0	10	5	1
4	0	1	10	6	1
4	0	1	10	6	1
5	1	1	10	6	3
5	2	3	10	7	5
5	2	3	11	9	5
5	2	3	12	10	5
5	2	4	15	10	6
5	3	4	15	10	7
5	3	4	16	10	9
5	3	5	20	11	10
5	3	5	20	13	10
7	4	5	20	15	10
8	5	5	20	15	10
10	8	6	21	16	10
10	9	8	21	18	21
10	11	10	21	20	
13	20	14	21	21	
15		15	21	21	
15			21		
20					
n = 32	n = 29	n = 30	n = 31	n = 30	n = 27

Notes:

WTP_P1: The WTP_P group's prediction of WTP_B

WTP_P2: The WTA_H group's prediction of WTP_B

WTP_P3: The WTA_B group's prediction of WTP_B

WTA_P1: The WTA_P group's prediction of WTA_B

WTA_P2: The WTP_H group's prediction of WTA_B

WTA_P3: The WTP_B group's prediction of WTA_B

Appendix 14: Difference between the Treatment Groups' Sample Statistics and the Sample Averages

Treatment	WTP_P	WTA_P	WTP_H	WTA_H	WTP_B	WTA_B
Average Age	1.154	-0.973	0.938	-1.579	0.780	-0.307
t-stat	0.687	-0.722	0.694	-0.891	0.436	-0.146
Male (%)	-0.020	0.060	0.028	0.028	-0.081	-0.022
t-stat	-0.220	0.646	0.302	0.297	-0.832	-0.235
Student (%)	-0.049	0.047	0.057	-0.002	-0.089	0.024
t-stat	-0.686	0.565	0.663	-0.019	-1.240	0.285
Income[a]	5.364	-1.919	-5.938	-0.751	1.053	1.979
t-stat	0.142	-0.046	-0.148	-0.018	0.026	0.048
Dependents[b]	-0.035	-0.102	0.156	-0.167	0.069	0.095
t-stat	-0.266	-0.594	1.207	-1.058	0.538	0.623
Education[c]	0.000	-0.016	0.086	-0.121	-0.056	0.100
t-stat	0.000	-0.174	0.928	-1.296	-0.562	1.083
Ethnicity[d]	0.014	-0.003	-0.076	0.131	0.138	**-0.188**
t-stat	0.150	-0.037	-0.827	1.388	1.407	**-2.256**
n	32	31	30	29	27	30

[a] Average weekly disposable income
[b] Average number of dependents
[c] Percentage with a degree or higher
[d] Percentage of European

Further notes:

1) Differences were calculated by subtracting the sample mean for each of the treatment groups from the sample mean for the experiment as a whole. For example, *Mean Age* in group WTP_P was 24.42 while the *Mean Age* for the experiment as a whole was 25.58. The difference is therefore 25.576-24.422 = 1.154.

2) t-statistics were calculated by dividing the difference between the two means by the standard error of the difference. The standard error of the difference was calculated using the formula $S.E(\mu m - \mu w) = \sqrt{(\frac{s.d_m^2}{n_m} + \frac{s.d_w^2}{n_w})}$ (Stock & Watson, 2007, p. 84).

3) A t-statistic less than -1.96 or greater than 1.96 indicates that the null hypothesis (difference between the two means is equal to zero) is rejected at a 5% level of significance.

Appendix 15: Standard Consumer Surplus Measure

WTP_B	Price	Consumer Surplus
$5.00	$4.20	$0.80
$5.00	$4.20	$0.80
$5.00	$4.20	$0.80
$5.00	$4.20	$0.80
$6.00	$4.20	$1.80
$6.00	$4.20	$1.80
$7.00	$4.20	$2.80
$8.00	$4.20	$3.80
$8.00	$4.20	$3.80
$10.00	$4.20	$5.80
$10.00	$4.20	$5.80
$12.00	$4.20	$7.80
$19.00	$4.20	$14.80
Total consumer surplus[a] =		$51.40

[a]Total consumer surplus equals the sum difference between what participants are willing to pay for organic cotton (WTP_B) and what they are required to pay (Price), counting only those participants whose WTP is greater than the price level (Mankiw, 2007).

Appendix 16: Mann-Whitney U Test – Ranking

WTA_B	Rank	WTP_B	Rank
1	8	0	2.5
1	8	0	2.5
1	8	0	2.5
1	8	0	2.5
6	24	1	8
7	27	1	8
7	27	1	8
8	30	2	13
10	35.5	2	13
10	35.5	2	13
10	35.5	3	16
10	35.5	3	16
10	35.5	3	16
10	35.5	4	18
12	40.5	5	20.5
15	43.5	5	20.5
15	43.5	5	20.5
15	43.5	5	20.5
15	43.5	6	24
16	46	6	24
18	47	7	27
20	49.5	8	30
20	49.5	8	30
21	54	10	35.5
21	54	10	35.5
21	54	12	40.5
21	54	19	48
21	54		
21	54		
21	54		
R_1	1137.5	R_2	515.5
n_1	30	n_1	27

Note: R_1 is the sum of the ranks for WTA_B data and R_2 is the sum of the ranks for the WTP_B data. If two or more WTA or WTP values are equal, then their ranks are summed and divided across each tied value.