University of St. Andrews

Discussion papers in Environmental Economics

http://www.st-andrews.ac.uk/gsd/research/envecon/eediscus/

Paper 2017-02

The External Validity of Consequential Stated Preference Studies: a Comment

D. A. Comerford and N. Hanley

Keywords: Cost Benefit Analysis; Stated Preference; Willingness-to-pay; Consequentiality; Act utility; Exogenous Goods; Preferences-over-actions; Preferences-over-outcomes

JEL codes: B41, C83, D61, D62, H41, H43

The External Validity of Consequential Stated Preference Studies: a comment

David A. Comerford[a*]

Nick Hanley[b]

Abstract

Mounting evidence suggests that Consequential Discrete Choice Experiments (CDCEs) are internally valid i.e. they elicit a de-facto revealed preference. This comment asks whether CDCEs are always externally valid. For instance, when it comes to existence values, policy makers require a valuation of the benefit that derives from *passively experiencing the continued existence of a good*, whereas CDCEs measure the value that derives from *actively intervening to maintain or increase the supply of a good*. We show that CDCEs will recommend suboptimal levels of Pigovian taxes and public goods provision. We suggest potential alternatives to CDCEs that future research should consider.

JEL Classification:

B41 Economic Methodology; C83 Survey Methods; D61 Cost Benefit Analysis; D62 Externalities; H41 Public Goods; H43 Policy Evaluation

Keywords: Cost Benefit Analysis; Stated Preference; Willingness-to-pay; Consequentiality; Act utility; Exogenous Goods; Preferences-over-actions; Preferences-over-outcomes.

[a*] Corresponding author: Division of Economics, University of Stirling, FK9 4LA, UK. email: david.comerford@stir.ac.uk

[b] School of Geography and Sustainable Development, University of St Andrews, KY16 9AL. email: ndh3@st-andrews.ac.uk. The paper benefited greatly from conversations with Leo Lades as well as with participants at the Stirling Behavioural Science seminar and 2016 Network of Intergrated Behavioural Science meeting at the University of East Anglia.

Introduction

Revealed preference has long been accepted by the economics discipline as the gold standard measure of the money-metric utility value of changes in goods and services (Samuelson, 1948: Beshears et al, 2008). Choices over non-market goods are often unobserved however. This missing market constraint characterises the task of valuing many public goods and externalities, particularly in the presence of passive-use values, which revealed preference non-market valuation methods such as travel costs and hedonic pricing are unable to measure, due to the lack of any suitable behavioural trail . To resolve this problem, the money-metric utility produced by changes in many non-market goods has been valued using stated preference surveys (Krutilla, 1967; Walsh, Loomis and Gillman, 1984. Stated preference surveys present respondents with a hypothetical market for a non-market good and elicit their willingness-to-pay for changes in the quantity or quality of that good in the context of this hypothetical market.

(figure 1 around here)

Figure 1 shows that from 2004 the phrase "stated preference" became more prevalent than the phrase "revealed preference". This should not be interpreted as signalling that stated preference has become as respected a measure of utility as revealed preference. Rather, it reflects the controversy that surrounds stated preference techniques; further indications of which include a US Supreme Court case from 2008 in which approximately $2bn was contested (Exxon Shipping Co. vs. Baker) and two separate symposia published in the Journal of Economic Perspectives from 2004 (vol. 8.4) and 2012 (vol 26.4).

This controversy has largely been concerned with the question: "do responses to stated preference surveys match respondents' revealed preferences?" This is a question of internal

validity, which is addressed in a recent survey paper of "best practice" in applying stated preference methods (Johnston et al, 2017). In the current research we consider a different concern: external validity. External validity refers to the generalizability of experimental results out of sample (Lynch, 1992). External validity is a core requirement of any preference elicitation procedure that seeks to estimate the costs and benefits that the population of final consumers derives from a good. In the current paper we ask: "do the preferences elicited in stated preference surveys match the preferences of the non-respondent population?".

1. Core Concepts

1.1. Consequential Discrete Choice Experiments (CDCE)

Carson and Groves (2007) provides the strongest evidence to date on the internal validity of stated preference measures. They demonstrate that stated preferences *are* revealed preferences if stated preferences are elicited under conditions of consequentiality in a single, binding referendum vote. Consequentiality requires that survey respondents believe that their answers "will influence agency decisions concerning the public good presented in the survey" (Carson, Groves and List, 2014, pp. 176-7), and/or will influence the cost that they will pay for an increase in public goods. In effect, consequentiality transforms a survey into a marketplace for public goods. The incentives that operate in a CDCE differ from those in a marketplace only with respect to probabilities; whereas decisions made in a marketplace typically achieve outcomes with certainty, decisions made in the CDCE will achieve outcomes with some positive probability. The value of consequentiality then is that it is incentive-compatible; it incentivises sincere and considered response. The recognition that CDCEs elicit revealed preference has been heralded as a paradigm shift in the valuation of non-market goods (Poe and Vossler, 2011).

Accumulating evidence suggests that CDCEs have demonstrated good internal validity. They correspond well with referendum results, which suggests that they eliminate the confounding factors that may have led stated preference to diverge from revealed preference (e.g. Herriges et al, 2010; Vossler, Doyon and Rondeau, 2013). They have been shown to comply with many of the criteria required of expected utility theory, as well as cumulative prospect theory and rank-dependent expected utility (Carson, Groves and List, 2014).

Carson, Groves and List (2014) set out five propositions to support their recommendation of consequentiality in preference elicitation. Proposition 5 states: "Replacing a vote of the entire population with an exogenously chosen random sample of the population does not alter the incentive properties of the mechanism." (Carson, Groves and List, 2014, p. 179). It is this proposition that we interrogate in what follows.

1.2.Preferences-over-outcomes versus preference-over-actions

In a critique of stated preference valuation, Hausman observes: "the neoclassical model requires that preferences be over states of the world and not over acts: for example, preferences must be over the choice between two different states of a wilderness area, not over whether the respondent receives a warm glow from the idea of saving a wilderness area" (Hausman, 2012, p.47). By definition, CDCEs elicit preferences over actions because consequentiality grants *agency* to the respondent. That is, consequentiality causes the respondent to believe that her response has some causal influence on the outcome that obtains. However, preferences-over-actions are conceptually distinct from preferences-over-outcomes, as set out in table 1.

(Table 1 around here)

Person *i*'s choiceless utility, $E[u_i(S_x)]$, refers to the utility *i* expects to derive from the state of the world that results from realization of outcome *x* if outcome *x* were brought about through some mechanism that is exogenous to person *i*. The term was coined by Loomes and Sugden, who define it as follows: "Choiceless utility is the utility that the individual would derive from the consequence *x* if he experienced it without having chosen it. For example, he might have been compelled to have *x* by natural forces, or *x* might have been imposed on him by a dictatorial government" (Loomes and Sugden, 1982). Choiceless utility is the sole source of utility that will inform a rational agent's preferences-over-outcomes when they have no agency, or think they have no agency.

In contrast, a rational agent's preferences-over-actions are informed by the sum of choiceless utility, $E[u_i(S_x)]$, and act utility, $E[u_i(A_X)]$. Person *i*'s act utility refers to

a) any utility that *i* derives from signalling endorsement of outcome *x*

and/or

b) any utility that *i* derives from being responsible for realization of outcome *x.*

Choiceless utility and act utility are modeled as separable because each is determined independently of the other: Choiceless utility is independent of how an outcome comes about whereas act utility is entirely determined by how an outcome comes about.

If person *i* considers herself to have some causal influence on the realization of outcome *x* person *i*'s act utility can be positive or negative. Examples of positive act utility that stem from acting generously are warm glow (e.g. Andreoni, 1990) and signalling status (e.g. Glazer and Konrad, 1996). Examples of negative act utility are the guilt or shame that accompanies acting meanly (e.g. Della Vigna et al, 2012) or the anticipated regret associated with setting a negative precedent (e.g. Spash and Hanley, 1995).

In section 3 we consider evidence that act utility has a substantive impact on choice in consequential DCEs. Importantly, CDCEs are sometimes used to estimate the value delivered by goods that cannot deliver act utility. The next subsection considers this type of good, which we term an exogenous good.

1.3.Exogenous goods

The defining feature of an exogenous good is that it is imposed on the consumer from without. As such, exogenous goods can deliver only choiceless utility. Unsolicited gifts are exogenous goods. All externalities are exogenous goods because consumers of externalities are, by definition, passive recipients of the good or bad outcomes that come their way (although of course they can take costly defensive or mitigating actions to reduce the harm they suffer from the externality). Environmental public goods such as landscapes and air quality are exogenous goods for all but the few consumers who believe themselves to have had a causal impact on their level of provision. Any policy *x* is also an exogenous good if it is paid for using tax revenue and if the taxpayer does not perceive herself as having had agency in bringing policy *x* into place.

CDCEs are often used to value exogenous goods (we discuss examples in section 5). The classic definitions of equivalent surplus and compensating surplus as welfare measures for a change in a public good indeed assume that the individual is not free to choose the level of supply of such public goods (and bads) – Johansson (1993). One reason why welfare economists measure preferences is to evaluate policy ex ante. Money-metric utility measures give policymakers an evidence base on which to make decisions. The motivating assumption behind compiling that evidence is that policymakers act as benign dictators: they seek to impose on the population public goods, policies and regulations that are expected to increase social welfare, according to the Kaldor-Hicks potential compensation test . The use of the

term “benign dictator” is deliberate; recall Loomes and Sugden’s examples of the circumstances when a consumer would experience only choiceless utility: “*x* might have been imposed on him by a dictatorial government” (1982). Whenever policy makers desire an ex-ante measure of the utility change that would be delivered by a policy that they would impose on the population, preference-over-outcome is the appropriate measure.

In the next section we present evidence that elicitations of welfare measures based on preference-over-action will deliver biased estimates of the utility engendered by exogenous goods.

2. Evidence of Act Utility

In this section we present evidence that act utility is a substantive source of utility that influences behaviour. An intuitive test for act utility is the question: would the final consumer derive the same utility from this outcome if it were exogenously imposed on them? If the answer to that question is “no”, then some act utility is at play.

One piece of evidence documenting act utility comes from a world-famous CDCE: the Brexit referendum. There is evidence that some people used their vote on the issue of the UK’s place in the EU to signal discontent on other issues; some even admitted that they regretted using the franchise as a protest (Dearden, 2016). Self-evidently these voters sacrificed choiceless utility (the expected utility of remaining in the EU) for the act utility that they derived from making a protest vote.

Experimental evidence of act utility comes from Della Vigna, List and Malmendier (2012). They estimate the scale of act utility in a situation that is often used as an analogue in explaining responses to a stated-preference survey for a pure public good: being solicited for a charitable donation (e.g. Kahneman and Ritov, 1993; Rondeau, Poe and Schulze, 2005). In

both CDCEs and charitable requests the agent is asked whether she is willing to make a monetary sacrifice in order to provide a service to others; in both situations, the service that is to be provided may be one with which the respondent is unfamiliar, or one that the respondent has not previously considered paying for. Della Vigna, List and Malmendier (2012) compare the likelihood of donating to charity when people are given advance warning that they will be asked for a donation versus when people are surprised by a request. If the only motivation for charitable giving is the choiceless utility that donors derive from helping charitable recipients, then the amount collected in the advance warning condition should be at least as high as the amount donated in the surprise condition. In fact, advance warning reduced donations by 28-42 percent. In other words, 28-42 percent of charitable donation in the surprise condition must have been driven by some motivation other than choiceless utility. That motivation is captured by our act utility term; it may be explained by social desirability bias, signalling, or avoidance of the negative act utility (guilt, shame) that is expected to result from not giving.

Note there are sources of act utility which are not accounted for in Della Vigna, List and Malmendier's 28-42 percent estimate. A notable source of act utility that will not be included in this estimate is the warm glow that people derive from donating (Andreoni, 1990). Della Vigna et al's result derives from subtracting one condition's donations from the other condition's donations. Since both conditions had the opportunity to give, some portion of giving by both groups might be explained by warm glow. Hence the 28-42% figure cannot include warm glow giving and cannot be interpreted as a comprehensive measure of act utility.

We next look to an incentive-compatible experiment that supplements Della Vigna et al (2012) in two important respects: it tests specifically for warm glow giving and it is conducted under conditions of anonymity. Crumpler and Grossman (2008) conducted an incentive-compatible laboratory experiment. Participants were asked which of ten charities they would like to receive $10 from the experimenters. Participants were then endowed with $10 and offered the opportunity to share some of their own $10 with their preferred charity. It was made clear to participants that the charity would receive a fixed payoff of $10, regardless whether the participant shared or not. The experimenters tested whether participants understood that any sharing would be exactly offset by a reduction in the donation to the charity. Despite the fact that participants knew that sharing would cost them money and leave the charity's payoff unchanged at $10, a majority of them still opted to share their endowment with the charity. They shared on average 20 percent of their endowment. This behaviour can only be explained by the presence of act utility.

We now look to a study that was explicitly designed to measure the magnitude of warm glow giving in a discrete choice survey experiment. Nunes and Schokkaert (2003) assume a constant level of warm glow derives from endorsing provision of any quantity of a moral good i.e. the warm glow *i* derives from supporting a higher quantity of moral good *x* is the same as the warm glow *i* derives from supporting a lower quantity of moral good *x.* If their assumption is correct then warm glow will cancel out when subtracting respondent *i*'s $wtp_{lower\ level}$ from respondent *i*'s $wtp_{higher\ level}$.[i] This allows retrieval of what Nunes and Schokkaert term a "cold" measure of WTP, one net of warm glow. Additionally, this method allows retrieval of WTP for warm glow. Nunes and Schokkaert's method estimates that between 35.5 and 73.5 percent of willingness to pay for the moral good is explained by warm glow.

Because Nunes and Schokkaert's analysis measures the bias induced by only one form of act utility, warm glow, it would be inappropriate to infer their estimate as a comprehensive measure of act utility, or even as a lower bound on act utility. Other results from the stated preference literature suggest that the positive act utility engendered by warm glow is sometimes offset by negative act utility. One example is a study that asked about a policy that would protect remnants of ancient Caledonian pine forest (Spash and Hanley, 1995). Over a quarter of the sample reported willingness to pay of zero, but in a debriefing every zero-bidder reported that they in fact derive positive choiceless utility from the continued existence of this ancient woodland. The explanation that accounted for over 70% of zero bids was that "biodiversity should be protected by the law". It seems these respondents reported a willingness to pay of zero in order to avoid the negative act utility of setting a precedent whereby market-like demand would decide whether such rare and threatened natural heritage survives or not. Protest responses of zero willingness to pay represent an instance where act utility depresses willingness to pay estimates.

A final example of act utility in a stated preference context comes from Frey and Oberholzer-Gee (1997). They conducted a referendum-style contingent valuation study which found that 50.8 percent of respondents supported a proposal to build a nuclear waste repository in their local area. A modified version of the survey proposed compensating households in the affected area thousands of dollars. This modification caused support to decline to 24%. The explanation for these backfiring incentives is that the introduction of compensation crowds out the motive to act like a good citizen. The choiceless utility engendered by the nuclear waste repository is clearly negative; otherwise the offer of it and thousands of dollars of compensation would be accepted. The act utility of helping society is so positive however that it offsets that negative choiceless utility. Frey and Oberholzer-Gee's stated preference

instrument was not consequential, but the results of the un-incentivized condition tally closely with the results of a referendum that was held a week later.

In summary, there is not yet a "gold standard" experiment on the magnitude of the bias induced by act utility on outcome-consequential WTP, but what we know is the following:

1. Incentive compatibility, or outcome-consequentiality, does not eliminate $wtp_{act\ utility}$ (e.g. Crumpler and Grossman, 2008; Della Vigna et al, 2012; Dearden, 2016).
2. Anonymity does not eliminate $wtp_{act\ utility}$ (e.g. Spash and Hanley, 1995; Nunes and Schokkaert, 2003; Crumpler and Grossman, 2008; Dearden, 2016).
3. Act utility in a DCE can be positive (Frey and Oberholzer-Gee, 1997; Nunes and Schokkaert, 2003) or negative (Spash and Hanley, 1995).
4. Estimates on the scale of bias induced by various sources of act utility are large: 28-42 percent (Della Vigna et al, 2012); 35 – 73 percent (Nunes and Schokkaert, 1998).

3. Implications of Act Utility for CDCEs

In the previous section we established that act utility is a substantive source of utility that is necessarily included in CDCEs. In this section, we consider the theoretical and practical implications of Act Utility for various tasks to which Discrete Choice Experiments are put.

3.1. Proposition 5 Invoked to Support Consequentiality Does Not Hold For Valuation of Exogeneous Goods.

Proposition 5 of Carson et al states: "Replacing a vote of the entire population with an exogenously chosen random sample of the population does not alter the incentive properties of the mechanism." (Carson, Groves and List, 2014, p. 179). Let's assume that what is at issue is a policy (henceforth policy x) that results in a change in the tax imposed on the

population. To see why proposition 5 does not always hold, let's consider three scenarios: a vote of the entire population; a vote of an exogenously chosen random sample; and a scenario in which the policy is imposed without a vote. In the scenario in which the policy is imposed without a vote, taxpayers are passive recipients of policy *x* and so can experience no act utility. In the scenario where there is a vote of the entire population on policy *x*, the entire population has the opportunity to experience act utility through their vote. If a voter derives warm glow from voting for the good cause that policy *x* represents to her, she can vote for the cause and enjoy that warm glow. The scenario where there is a vote of an exogenously chosen random sample grants the opportunity to experience act utility only to those taxpayers who are randomly selected to participate in the CDCE; non-respondents remain passive recipients of whatever outcome is selected by respondents and so experience no act utility. Since act utility is a source of substantive (positive or negative) utility when responding to CDCEs, respondents receive substantively different utility than non-respondents. Replacing a vote of the entire population with an exogenously chosen random sample of the population *does* substantively alter the incentive properties of the mechanism.

3.2. Modelling the Bias that results from using CDCE's to Estimate the Benefits Provided by Exogenous Goods

The total population-level benefits that the introduction of exogenous good x delivers can be represented as in equation (1):

$$\text{benefit} = \theta * wtp_r + (1 - \theta) * wtp_n \qquad (1)$$

Where wtp_r is a money measure of *preference-over-action* for bringing the exogenous good into being; wtp_n is a money measure of non-respondents' *preference-over-outcome*; and θ is the proportion of respondents in the population of final consumers.

The scale of the bias that results from using CDCE's to estimate benefits is modelled in equation (2). Equation (2) is simply the benefit estimate for final consumers, as stated in equation (1) above, subtracted from the benefit estimate implied by respondents, wtp_r.

Bias = $wtp_r - [\theta * wtp_r + (1 - \theta) * wtp_n]$ (2)

Equation (2) reduces to:

Bias = $(1 - \theta) * (wtp_r - wtp_n)$

Assuming respondents are truly selected at random, then the only difference between wtp_r and wtp_n is act utility.

By assumption, $wtp_r - wtp_n = wtp_{act\ utility}$

Hence bias reduces to two parameters: bias increases with the scale of act utility, $wtp_{act\ utility}$, and decreases with the sample size of the discrete choice survey experiment, θ. As we saw in the previous section, the available evidence suggests $wtp_{act\ utility}$ is large. Sample size considerations have not warranted much attention in the stated preference literature, probably because the prevailing wisdom has been that a representative sample would reveal utility measures that are representative of the utility enjoyed by the population of final consumers (e.g. proposition 5, Carson, Groves and List, 2014). Where sample size has come in for discussion, it has been with regards to reliability e.g. on the basis of the tightness of confidence intervals. DCEs generally recruit samples that are well below 1 percent of the population of final consumers. Typically then, even a very small $wtp_{act\ utility}$ can scale up to produce a large bias at the aggregate value level.

4. For What Tasks is Consequential DCE Externally Valid (and for what tasks is it not)?

We consider the following tasks to which DCEs have been put:

- Predicting the results of a Referendum on Public Good Provision

- Conducting Cost-Benefit Analysis of Provision of a Public Good
- Determining Optimal Pigovian Tax/ Subsidy Levels
- Predicting Demand for a Proposed Private Good

The criterion for external validity is that non-respondents value the good similarly to respondents i.e. $wtp_r = wtp_n$ (Lynch, 1992). Absenting the case where act utility happens to be zero, a necessary condition in order that $wtp_r = wtp_n$ is that non-respondents can experience act utility. It is this condition that we test in what follows.

4.1. Predicting results of a referendum

In some societies (e.g. Switzerland, Colorado), local public good provision is at least partly determined by referenda. A referendum is effectively a CDCE: voting in favour of a policy makes that policy more likely to result, voting against it makes it less likely. Respondents in a referendum can derive act utility from that agency, as pointed out in the discussion of the Brexit referendum in section 3. Since both the population and the sample of respondents to the CDCE enjoy act utility, CDCEs may be externally valid for predicting the results of a referendum. Indeed, empirical comparisons suggest that CDCEs can predict the results of referenda well (e.g. Herriges et al, 2010; Carson, Groves and List, 2014).

Result 1: CDCEs may be externally valid for predicting voting in a referendum.

4.2.Valuing a Public Good in the context of cost-benefit analysis

A CDCE measures act utility that an exogenous good cannot deliver. Since public goods are exogenous goods, $wtp_r \neq wtp_n$.

We will show that an outcome-consequential stated-preference study can recommend provision of a public good that is certain ex-ante to deliver a deadweight loss i.e. is certain to result in a net welfare loss. Symmetrically, a CDCE can reject provision of an exogenous good that is certain to deliver a net welfare benefit.

Take the case where an increase in provision of an exogenous good *x* (e.g. species preservation) is proposed. The cost to each household of providing good *x* is estimated at £8. At $t = 0$, a CDCE is conducted on a sample of respondents, who are randomly selected and are representative of the population. The study finds that households are, on average, willing to pay £9 for public good *x*. The CDCE implies then that gains (£9 per household) exceed losses (£8 per household) and hence that provision of good *x* satisfies the Kaldor-Hicks criterion. Consequentiality would be expected to deliver this good.

For the sake of this example, let us assume that £3 of the £9 willingness to pay figure derives from act utility[ii]. The remaining £6 of the £9 willingness to pay is a measure of the choiceless utility derived from final consumption of the good.

Now let us imagine that it is $t = 1$ and good *x* has been provided. We conduct a survey on another random, representative sample of households drawn from the same population as was the $t = 0$ study. Assume that this survey can accurately elicit an honest response to the question: "*how much is good* x *worth to you?*" Let's further assume that at $t = 0$ respondents to the outcome-consequential survey accurately anticipated benefits and honestly reported their WTPs.

At $t = 1$ respondents to the outcome-consequential study will report that good *x* is worth £9 to them; £6 of which is the utility they derive from that state of the world achieved by good *x*, and £3 of which is the utility they derive from having been a causal agent in providing good *x*. The remainder of the population, who were not respondents to the outcome-consequential

stated-preference study, will report that good *x* is worth £6 to them, since they do not receive warm glow or any other form of act utility from the exogenous provision of good *x*.

The population-level benefits that public policy *x* has delivered is measured by equation (1), which we encountered in subsection 4.2.

benefit = $\theta * wtp_r + (1 - \theta) * wtp_n$ (1)

Let's assume that this study followed Carson's sample size recommendation of "a couple thousand" (1996), and the relevant population is a couple of hundred thousand, so that θ is 1% of the population of final consumers. Then, the benefit derived from good *x* by the average member of the population is given by:

$benefit_x$ = (.01)*(\$9) + (1 - .01)*(\$6) = \$6.03

In this case, the estimated benefit for the average household is less than the estimated \$8 cost for the average household, and so good *x* does not satisfy the Kaldor-Hicks criterion. In summary, a CDCE would be expected to recommend a provision of a welfare-losing public good.

A CDCE could also lead to rejection of a welfare-enhancing public good. All that is required is that the act utility engendered by endorsing the good is negative e.g. that it elicits protest responses of zero (see Spash and Hanley, 1995). Let's retain the above example, but imagine that costs are now \$5 and act utility is now -\$2. The average measure of benefit derived from good *x* enjoyed by the average member of the population is given by:

$benefit_x$ = (.01)*(\$6-\$2) + (1 - .01)*(\$6) = \$5.98

The utility expected for the average final consumer \$5.98 is greater than the \$5 cost, and so with appropriate measures of benefits the Kaldor-Hicks criterion would recommend provision of this good. However, the outcome-consequential measure is inappropriate because it

implies that all consumers will realize benefits of just $4. The ultimate evidence that the CDCE is not externally valid is that it will result in rejection of a welfare-enhancing good.

Result 2: Consequential DCEs are not externally valid for measuring the utility engendered by public goods in the context of cost-benefit analysis.

4.3. Setting Pigovian Taxes

Pigovian taxes are imposed on consumption or production that inflicts costs on third parties e.g. smoking inflicts health costs on bystanders. Pigovian taxes will ideally be set so as to exactly match the marginal external cost that consumption inflicts on third parties at the socially-optimal level of output (Hanley et al, 2007). When revealed preference measures of marginal external cost are not available, stated preferences are often elicited to put a monetary estimate on those costs (e.g. Scarpa, Willis and Ocutt, 2007; Longo, Markandya and Petrucci, 2008; Aravena, Hutchinson, and Longo, 2012). Pollution as an external cost is an exogenous public bad and so cannot deliver act utility to the final consumer, whereas the CDCE will include act utility in its measure, with the result that $wtp_r \neq wtp_n$. As with the public goods example in 5.2., if the act utility of mitigating external costs is positive (negative) then respondents will tend to overstate (understate) marginal external costs, and too large (small) a tax will be set, leading to underconsumption (overconsumption).

Result 3: Consequential DCEs are not externally valid for setting Pigovian taxes.

5. How to value non-market exogenous goods?

The contribution of this comment is to make the case that CDCE procedures will be expected to yield biased estimates of the money-metric utility delivered by exogenous goods when act utility figures in people's utility functions. Society will misallocate resources if exogenous

non-market goods, such as species protection, are systematically under- or over-valued. What is to be done to de-bias these valuations? Can the desirable features of consequentiality – that it encourages people to carefully consider their choices as a constrained optimization problem – be retained, without incurring such bias? Here we offer some suggestions.

In theory, one solution to this problem is to grant everyone in the population agency over outcomes by holding a referendum of the population whenever we wish to value a public good or estimate marginal external costs. In practice, this approach is prohibitively burdensome. Additionally, a state's legitimacy would be compromised if the results of referenda were not implemented as policy. In order to maintain its legitimacy, a state would have to abandon the welfare criterion in favour of a popular vote.

A more promising solution is the use of subjective wellbeing data. Subjective wellbeing data reveals a measure of the utility that members of a population enjoy from actual receipt of a good. It measures experienced responses to realized outcomes. Those responses will include act utility if the respondent derives utility from her causal influence in bringing about an outcome (see Frey, Benz & Stutzer, 2003, for a discussion and Botti & McGill, 2011, for an empirical example). If the respondent had no causal role in bringing about an outcome, then her implied valuation of a good will measure choiceless utility only. With a representative sample of the population, a survey measure of subjective wellbeing will appropriately weight preference-over-outcomes and preference-over-actions to deliver a population-level estimate of the total utility a population derives from a good.

On the face of it, responses to subjective wellbeing questions do not appear incentive compatible – there is no obvious cost to a respondent who chooses to misrepresent how she feels. On the other hand, for the valuation of public goods, subjective wellbeing questions do not lend themselves to strategic response, because a respondent cannot know at the time of

response which way to bias her answer so as to increase valuation of an outcome she desires. The dominant strategy of a respondent who knows that subjective wellbeing responses will be used to value some as yet unspecified public goods or decide on their supply is to answer honestly – an honest answer will at least imply positive valuation of those goods that she experiences as beneficial, and imply negative valuation of those goods that she experiences as costly. As a means to elicit unbiased utility measures, subjective wellbeing questions therefore seem to have similar incentive properties to CDCEs.

As things stand, Subjective Wellbeing data could substitute for stated preference in only a very narrow subset of the valuation tasks to which stated preference surveys are put. Stated preferences are often used to measure the utility that would be engendered by some, as yet, unavailable good. There can be no experienced utility measure of goods that have not yet been experienced. Even when we wish to measure the utility engendered by the continued existence of a public good e.g. a threatened ecosystem (e.g. Spash and Hanley, 1994), we will not have the requisite data unless a survey of subjective wellbeing has been conducted on a sufficiently large sample and with sufficient exogenous variation in receipt of the good as to provide causal identification of the utility engendered by the good. If these limiting conditions do not hold then a stated preference survey will be required. A vital question for future research is what form that survey should take.

A third means to value the utility delivered by exogenous goods is through the use of non-consequential, non-choice, stated preference methods. Instead of asking respondents "which do you choose?", we could ask them "which outcome would you prefer to receive?" The linguistic distinction is subtle, but the latter question is devoid of agency and hence should measure preference-over-outcome only. It remains to be seen whether respondents attend to

its linguistic nuance. Also, this question abandons incentive-compatibility, which re-introduces the possibility of hypothetical bias. Some will dismiss the proposed question for this reason alone. On the other hand, we have suggested here that CDCEs elicit preferences that are not externally valid when act utility is present. Future researchers face a choice between a technique that delivers an accurate measure of the wrong preference or a technique that delivers a biased estimate of a more appropriate welfare measure. On this point, it is worth noting that the most recent and comprehensive meta-analysis of hypothetical bias in the literature found that the median level of hypothetical bias across 83 observations was 35 percent (Murphy et al, 2005). The scale of bias induced by sources of act utility was 28-42 percent (Della Vigna et al, 2012) and 35 – 73 percent (Nunes and Schokkaert, 2003).

6. Discussion

Environmental public goods – particularly those which generate substantial passive use or existence values – have been vulnerable to damage, due to perceived methodological weaknesses of stated preference, the dominant methodology for measuring their value. Stated preferences have been argued to lack the incentive-compatibility of revealed preference. The internal validity of stated preferences took an enormous step forward with the publication of Carson and Groves (2007). That paper sets out clear conditions under which stated preferences are de facto revealed preferences. It is tempting to think that there is now a tool that can be applied off-the-peg to deliver unbiased valuations of prospective changes in the supply of environmental public goods. The contribution of the current research is to demonstrate why we are not yet at that point.

The question CDCEs ask ("*which outcome will you bring into effect?*") does not yield an appropriate measure of the benefits delivered by goods that the population passively receives.

CDCEs will misrepresent the costs and benefits that non-respondents derive from exogenous goods.

There is a nuance to our argument that warrants explication and emphasis at this point. Though a CDCE is expected to lack external validity, a population-wide referendum that asks an identical question is not. A population-wide referendum grants act utility to both the respondents and the affected population, whereas a CDCE grants act utility to the respondents only.

Here we have been concerned only with the role of act utility ex ante, i.e. how act utility impacts on choice. There is a literature on ex-post act utility that we have not considered here, but the interested reader should look to Frey, Benz and Stutzer (2003) for an excellent review. A result that consistently is demonstrated is that subjects (employees, taxpayers) are more satisfied with outcomes when they perceive themselves as having agency over those outcomes (e.g. Spencer, 1986; Frey and Stutzer, 2005).

For the purposes of this paper we have set aside the important question of whether ex-ante act utility aligns with ex-post act utility. It is possible that choosers make systematic mistakes when predicting the stream of benefits delivered by act utility. We did touch on some evidence documenting a gap between predicted act utility and experienced act utility: voters in the UK expressed regret about using the Brexit referendum as a protest vote (Dearden, 2016). Investigating this decision – experience gap is an important area for future research on act utility but we have set it aside in this paper because, as regards theory, it is a separate source of bias to the external validity bias which concerns us here. Whereas the problem we have outlined is a discrepancy across respondents and non-respondents, the decision - experience gap occurs within a given respondent such that respondent *i*'s prediction of $E[u_i(S_x)]$ is subject to systematic cognitive biases that lead to divergence from respondent *i*'s

realized $u_i(S_x)$. The decision – experience gap exacerbates the external validity bias discussed throughout this paper.

The external validity bias that is the focus of our research does not occur within the respondent; it is a consistent measurement error on the part of the researcher. Researchers are making a category mistake when they take respondents' preferences-over-action as representative of non-respondents' preferences-over-outcome. A key question is how to correct this category mistake.

We suggested three potential methods. One solution to this problem is to convert every passive non-respondent into an agentic respondent by holding a referendum of the population whenever we wish to value a public good or estimate marginal external costs. This approach opens up other problems however: it is extremely burdensome, prohibitively so, in our opinion. A second approach is to use experienced utility measures such as subjective wellbeing data derived from a representative sample of the population, since these data will necessarily be representative of the outcome utility of the population of final consumers. A third approach is to measure preference-over-outcome by asking respondents to a stated preference survey to report their preference-over-outcome. This would require removing the respondents' agency from the preference elicitation question. So, instead of asking "which do you choose?" the survey question would ask "which would you prefer to receive?" Adopting this third approach measures the appropriate form of utility, but of necessity, it abandons consequentiality. We look forward to seeing the development of preference measures that are both externally- and internally-valid, and which can resolve the apparent tension which we highlight here.

References:

Andreoni, James. 1990. “Impure altruism and donations to public goods: A theory of warm-glow giving”. *The Economic Journal*, 100 (401): 464-477.

Aravena, Claudia, W. George Hutchinson, and Alberto Longo. 2012. "Environmental pricing of externalities from different sources of electricity generation in Chile." *Energy Economics* 34 (4): 1214-1225.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2014. "Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices." *The American Economic Review* 104, (11): 3498-3528.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "How are preferences revealed?." *Journal of Public Economics* 92, (8): 1787-1794.

Botti, Simona, and Ann L. McGill. "When choosing is not deciding: The effect of perceived responsibility on satisfaction." Journal of Consumer Research 33, no. 2 (2006): 211-219.

Carson, Richard.T., 2000. Contingent valuation: a user's guide. *Environmental Science & Technology*, 34: 1413-1418.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods." *Land Economics* 72 (1): 80-99.

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

Carson, Richard T., Theodore Groves, and John A. List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists* 1: 171-207.

Crumpler, Heidi, and Philip J. Grossman. 2008. An experimental test of warm glow giving. *Journal of Public Economics* 92 (5-6): 1011-1021.

Dearden, Lizzie. 2016. Anger Over Bregret as Leave Voters Say They Thought UK Would Stay in EU. Independent. http://www.independent.co.uk/news/uk/politics/brexit-anger-bregret-leave-voters-protest-vote-thought-uk-stay-in-eu-remain-win-a7102516.html

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics* 127: 1-56.

Exxon Shipping Co v Baker, 2008.128 Supreme Court 2605

Frey, Bruno S., Matthias Benz, and Alois Stutzer. "Introducing procedural utility: Not only what, but also how matters." Journal of Institutional and Theoretical Economics JITE 160, no. 3 (2004): 377-401.

Frey, Bruno S., and Felix Oberholzer-Gee. "The cost of price incentives: An empirical analysis of motivation crowding-out." The American economic review 87, no. 4 (1997): 746-755.

Frey, Bruno S., and Alois Stutzer. 2005. "Beyond outcomes: measuring procedural utility." *Oxford Economic Papers* 57 (1): 90-111.

Glazer, Amishai., and Kai. Konrad. 1996. “A signaling explanation for private charity.,” *American Economic Review*, 86 (4): 1019–1028.

Hausman, Jerry, 2012. Contingent valuation: from dubious to hopeless. *The Journal of Economic Perspectives*, 26 (4): 43-56.

Herriges, Joseph, Catherine Kling, Chih-Chen Liu, and Justin Tobias. "What are the Consequences of Consequentiality?." Journal of Environmental Economics and Management 59, no. 1 (2010): 67-81.

Johansson, Per-Olov (1993) *Cost-Benefit Analysis of Environmental Change*. Cambridge University Press.

Johnston, Robert J., Kevin J. Boyle, Wiktor (Vic) Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau, and Christian A. Vossler (2017) Contemporary Guidance for Stated Preference Studies *Journal of the Association of Environmental and Resource Economists.* In press.

Kahneman, Daniel, and Ilana Ritov. 1994. Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty*, 9: 5-37.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of experienced utility." *The Quarterly Journal of Economics* 112 (2): 375-405.

Krutilla, John. 1967. Conservation reconsidered. *American Economic Review* 57: 787 –796.

Longo, A., Markandya, A. and Petrucci, M., 2008. The internalization of externalities in the production of electricity: willingness to pay for the attributes of a policy for renewable energy. *Ecological Economics*, 67 (1): 140-152.

Loomes, Graham, and Robert Sugden. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92, no (368): 805-824.

Luechinger, Simon. 2009. Valuing air quality using the life satisfaction approach. *The Economic Journal* 119, (536): 482-515.

Lynch, John. 1982. On the External Validity of Experiments in Consumer Research. *Journal of Consumer Research* 9: 225–239.

Nunes, Paulo. A.L.D. and Erik Schokkaert. 2003. Identifying the warm glow effect in contingent valuation. *Journal of Environmental Economics and Management* 45: 231-245.

Poe, Gregory L., and Christian A. Vossler. 2011Consequentiality and contingent values: an emerging paradigm. Edward Elgar Publishing: Cheltenham, UK.

Reinstein, David, and Joon Song. 2012. "Efficient consumer altruism and fair trade products." *Journal of Economics & Management Strategy* 21 (1): 213-241.

Rondeau, Daniel., Gregory. L. Poe and William. D. Schulze. 2005. VCM or PPM? A comparison of the performance of two voluntary public goods mechanisms. *Journal of Public Economics*, 89: 1581-1592.

Samuelson, Paul A. 1948. Consumption theory in terms of revealed preference. *Economica*. 15 (60): 243–253.

Scarpa, R., Willis, K.G. and Acutt, M., 2007. Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management*, 50 (4). 449-466.

Spash, Clive L., and Nick Hanley. 1995. Preferences, information and biodiversity preservation. *Ecological Economics* 12: 191-208.

Spencer, Daniel G. 1986. "Employee voice and employee retention." *Academy of Management* 29 (3): 488-502.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4: 145-171.

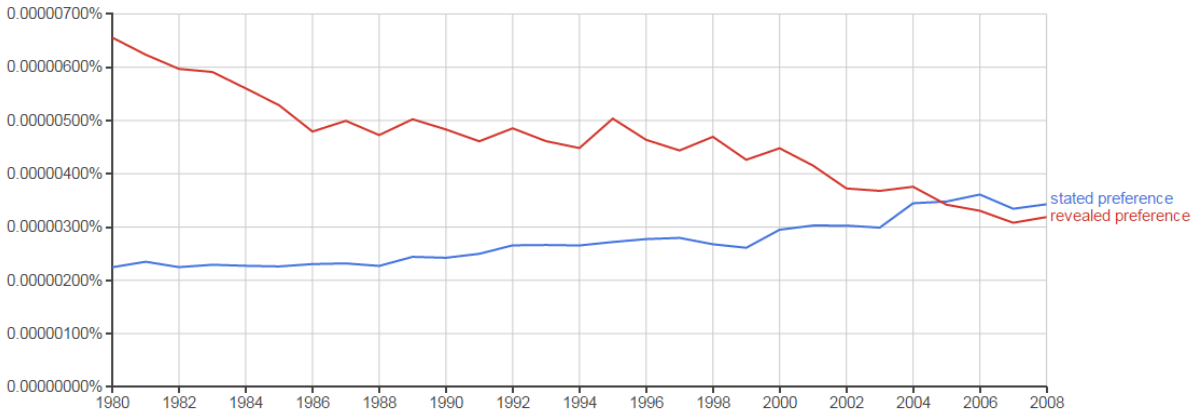
Vossler, Christian A., Joe Kerkvliet, Stephen Polasky, and Olesya Gainutdinova. 2003. "Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon." *Journal of Economic Behavior and Organization*, 51(2): 261–77.

Walsh, Richard G., John Loomis, Richard A. Gillman (1984) Valuing option, existence and bequest demands for wilderness *Land Economics*, 60, 14-29.

Table 1: A Framework for Distinguishing Categories of Preference and Utility

Type of preference	Informed by	Informed by utility function	Defining characteristic
Preference-over-outcome	Choiceless utility only	$E[u_i(S_x)]$	Person *i* has no agency over outcome *x*
Preference-over-action	Choiceless utility and act utility	$E[u_i(S_x)] + E[u_i(A_x)]$	Person *i* has agency over outcome *x*

Figure 1: Prevalence of reference to “stated preference” (red) and “revealed preference” (blue) in books published in American English, 1980-2008.



Footnotes:

[i] Note that if their assumption is incorrect, then the alternatives are that warm glow decreases with the scale of the public good or warm glow increases with the scale of the public good. The former alternative is implausible, and if the latter is true then Nunes and Schokkaert's approach will understate $wtp_{warm\ glow}$.

[ii] This assumption is not unreasonable on the basis of empirical estimates of WTP for sources of act utility, see section 3.