Cost Benefit Rules when Nature Counts

Olof Johansson-Stenman

Working papers in Economics no. 198, this version 2006-05-09
Department of Economics, School of Business, Economics and Law,
Göteborg University, Sweden

Abstract This paper analyses normative implications of relaxing the conventional welfare economics assumptions anthropocentrism and welfarism, i.e. that only human well-being counts intrinsically, combined with various types of non-selfish individual preferences. Social decision rules are derived for the optimum provision of a public good (environmental quality). It is shown that in several cases analysed, the basic Samuelson rule still holds, in terms of aggregate marginal willingness to pay.

Keywords: Altruism, welfarism, anthropocentrism, cost-benefit analysis, public good provision, social preferences, conditional cooperation

JEL: D6, D7, Q5

Acknowledgement: I am grateful for constructive comments from seminar participants at the London School of Economics and Political Science and the Stockholm School of Economics. Financial support from the Swedish Research Council is gratefully acknowledged.

Ethical forces are among those of which the economist has to take account. Attempts have indeed been made to construct an abstract science with regard to the actions of an "economic man," who is under no ethical influences and who pursues pecuniary gain warily and energetically, but mechanically and selfishly. But they have not been successful, nor even thoroughly carried out.

Alfred Marshall (1890), Preface to *Principles of Economics*, first edition.

1. Introduction

What is intrinsically valuable from a social point of view? This is not a question for which it is easy to provide a clear answer with which most people would agree; indeed, it is an issue with which philosophers have been struggling for centuries. Still, within welfare economics there is less heterogeneity. Here policy recommendations are most often based, explicitly or implicitly, on a Bergson-Samuelson social welfare function $W = w(U^1, U^2, ..., U^n)$, where U^k represents utility (a measure of well-being or preference satisfaction) of person *k*, and where *w* is increasing in its arguments. The extent varies to which further assumptions are made with respect to the specific structure of *w* (e.g. utilitarianism) and of utility measurability (e.g. along ordinal versus various kinds of cardinal scales) as well as interpersonal comparability of utilities. In some instances one can conclude that $W(\text{Policy } A) > W(\text{Policy } B)$ without any further assumptions and to argue from this relationship that Policy *A* is preferred to Policy *B*. That more social welfare is better than less is in this case equivalent to saying that a Pareto improvement is good.[1] Within welfare economics, this is typically seen as a weak and ethically uncontroversial condition for policy.

However, critics remain, of whom Amartya Sen is perhaps the most well-known. He has repeatedly argued against the underlying *welfarism* assumption in welfare economics, i.e. that social welfare depends solely on individual utility (Sen 1970, 1979). According to him,

[1] This was, for example, the formulation of Pareto efficiency used by Samuelson (1954) to derive the condition for Pareto efficient public good provision, subsequently known as the Samuelson rule.

there are other aspects that should be valued *intrinsically*, and hence not solely *instrumentally* through the utilities, including various "functionings", freedom, and some basic rights (e.g. Sen 1979, 1985a, 1993). If so, it is not at all clear that a Pareto improvement is good. Even if utility does not decrease for anyone, some other negative consequence may arise, such as reduced freedom. If so, one has to compare the utility improvement (for someone) with the freedom reduction (possibly for someone else) before such a conclusion can be made.[2]

This paper will formally analyse the consequences of relaxing the welfarism assumption, but based on a different motivation from Sen's, namely that there is evidence that at least some people value the environment *per se*, i.e. irrespective of the utility that the environment brings us. Second, it will analyse welfaristic but non-anthropocentric SWFs, i.e. it will allow for animal well-being to count directly in the SWF.

Of course, all models are simplifications of reality, and as such they can always be generalised and less restrictive, and hence in a sense be more realistic. However, the benefits of increased realism do not necessarily outweigh the costs associated with more complicated models that for example typically imply more ambiguous results. Even so, an accompanying paper to this one (Johansson-Stenman, 2006) argues that there are good reasons to believe that the potential benefits from relaxing the assumptions are sometimes large. Three main reasons were presented: *i*) The discussion within moral philosophy of animal suffering provides little support for anthropocentrism. *ii*) Evidence from environmental valuations studies seem to indicate that a substantial share of the respondents have non-anthropocentric and/or non-welfaristic preferences. *iii*) Evidence was presented from a large representative survey in Sweden where the respondents were explicitly asked about their ethical perceptions. It turned out that a majority have ethical preferences that are broadly consistent with consequentialism, but that very few have ethical preferences consistent with

[2] Correspondingly, this implies that any non-welfarist method for evaluation will in general imply that the

anthropocentrism.

This paper combines welfare economic models, where the welfarism and anthropocentrism assumptions are relaxed to varying degrees, with insights from modern behavioural economics, in particular the social preference literature where it is found that people often deviate substantially and systematically from the narrow *Homo Economicus* predictions. Policy implications with respect to the optimal provision of environmental (or more generally publicly provided) goods are derived in a two-individual economy for a number of cases. Section 2 presents the standard model based on anthropocentric welfarism, where people's utility functions are characterised by pure utility-oriented altruism as well as paternalistic environment-focused altruism. The novelty in this section is that this model is extended with the assumption that people's utility also depends on animal suffering and environmental quality directly. It is shown that, irrespective of these altruistic concerns and utility interdependencies, the basic cost-benefit rule (the Samuelson 1954 rule) for improved environmental quality still implies a Pareto efficient allocation, provided that each individual knows that the other will, on the margin, pay his/her maximum *WTP* for increased environmental quality.

Section 3 keeps both the social objective and the individual utility functions reflecting individual well-being, but at the same time assumes that people do not respond to *WTP* questions as utility-maximising consumers anymore, but as social welfare maximizing "citizens" as suggested e.g. by philosopher Mark Sagoff (1988, 2004). However, despite this it is, perhaps surprisingly, shown that the basic cost-benefit rule is still appropriate to use (again as long as each individual pays his/her maximum WTP on the margin for the good). Section 4 broadens the perspective further by allowing non-welfaristic and non-anthropocentric social welfare functions. Here, not surprisingly, the optimality results tend to

optimum is not Pareto efficient; see Kaplow and Shavell (2001).

be somewhat more complicated. However, if at least one of the individuals responds to the WTP questions as a social welfare maximising citizen, the standard cost-benefit rule still implies social welfare maximisation (which in this case is in general not Pareto efficient) if the arising co-ordination problem could be solved. However, if instead people respond as utility-maximising consumers, the standard cost-benefit rule would imply an under-supply of environmental quality. Overall, it is interesting that the basic Samuelson rule continues to hold in many cases involving several kinds of altruism and interdependent preferences as well as non-anthropocentric ethical assumptions. Section 5 concludes the paper.

2. The Welfaristic and Anthropocentric Model

For analytical convenience we focus on a two-individual economy,[3] consisting of Alice and Bob, who are both friends of each other and environmentalists. They both derive utility from private income x and environmental quality E. We allow for both pure utility-based (or non-paternalistic) altruism and paternalistic (in our case environment-focused) altruism. Thus, Alice derives utility from Bob's utility and in addition to this she may also derive utility from Bob's environmental quality *per se* (and vice versa). The latter implies that for constant levels of Alice's perceived environmental quality and Bob's utility, Alice's happiness increases if Bob's perceived environmental quality improves.[4] There are several papers that allow for both of these types of altruism (e.g. Archibald and Donaldson 1976; Jones-Lee 1990; McConnell 1997; Bergstrom 1999). In this paper we extend this to allow for

[3] In principle, all results can be generalised for a many-individual economy.

[4] See Jacobsson et al. (2006) for recent empirical evidence that people's altruism is often paternalistic in nature, and Johansson-Stenman (2005) for evidence that, whether the altruism is paternalistic or not, it can have large implications for behaviour such as the extent to which rich countries choose to internalise the environmental costs that they impose on poor countries.

preferences directed towards animal well-being *A*, where $A = \psi(v^1, v^2, ..., v^n)$ is an index variable reflecting animal well-being which is negatively related to the suffering of each animal v^i. Then we have:

$$U^A = u^A(x^A, E, A(E), U^B, E^B(E)) \quad (1)$$

where the utility function is increasing and quasi concave in its arguments (to ensure a unique optimum). Consider now a standard contingent valuation (CV) maximum willingness to pay (WTP) question as follows: *what is the maximum amount that you would be willing to pay for a certain (small) increase in E, provided that others would also have to pay a corresponding amount?* The phrase "the corresponding amount" is often more explicitly expressed as, for example, "the same amount as you," or "proportional to their after-tax income" etc. For analytical convenience we will here follow e.g. Bergstrom (2006) and focus on the situation where others (in this case either Alice or Bob) are paying their maximum WTP. Moreover, we will focus throughout on small, continuous changes so that we can disregard income effects,[5] although, the main conclusions carry over to the discrete case as well. For simplicity we will also assume a first-best economy in the sense that no distortionary taxes are needed in order to raise public revenues.[6] Finally, and most importantly, we will assume that people

[5] See e.g. Bergstrom (2005) for a recent analysis that models such income effects explicitly.

[6] This assumption is standard in the literature on interdependent preferences and altruism. Moreover, the value added from modelling distortionary taxes in an economy with identical individuals where lump-sum taxes cannot be used is dubious, since such second-best models do not rely on realistic information limitations because there are no problems *per se* with using head-taxes. It has been shown that insights from such models can be very misleading compared with the more realistic set-up with exogenous differences in ability and the possibility of using non-linear income taxation which implicitly allows for uniform (but not differentiated) lump-sum taxes; see e.g. Boadway and Keen (1993). One alternative would be to use a Stiglitz-type of self-selection model where the problem is to choose the optimal amount of public good to provide simultaneously as choosing tax parameters for the optimal non-linear tax problem. However, such an exercise would presumably

answer these types of question honestly (as is typically assumed in the environmental valuation literature) given their preferences, which are assumed to be perfectly known to the respondents themselves. Moreover, we assume that the responses do not reflect either "warm glow" (Andreoni 1989, 1990) or "purchase of moral satisfaction" (Kahneman and Knetch, 1992) that has been gained by the respondents from the elicitation process.[7] Whether these are realistic assumptions or not is heavily debated; see e.g. Diamond and Hausman (1994) for a critique of such methods. However, the purpose of this paper is not to analyse or discuss the value elicitation methods *per se*. Irrespective of how reliable these methods are, it is nevertheless important to analyse the welfare implications as though we could observe honest responses, and this is the task pursued here.

Similarly, if instead respondents are asked a so-called referendum-type dichotomous choice CV question such as "would you be willing to pay Y Dollars for a certain small improvement, conditional on the fact that others would also pay a corresponding amount?", which is the most frequently used format after the NOAA-panel's (Arrow et al. 1993) recommendation, they are instead assumed to be indifferent when their utility is held constant. In such a situation, it can be shown (see appendix for proofs) that Alice's *MWTP* for

---

provide fairly limited insights except for some special cases that rely on very strong separability assumptions.

[7] It has been heavily debated whether the welfare associated with the "warm glow" from contributing to a good social cause should be included in social welfare analysis or not. For example, Diamond and Hausman (1994) argued that such values should not be included whereas Harrison (1992) argued that the respondents' *motives* are irrelevant: "I call my utility 'jolly'. What you call your utility is ... your business" (Harrison 1992, p. 150). The position taken here, following Johansson-Stenman (1998, 2002), is that warm glow feelings are as real as other feelings, and that there is no reason to exclude such welfare contributions *per se*. However, the purpose of cost benefit analysis is not to estimate the respondents' instantaneous welfare *from responding to CV questions*, but rather to elicit responses that are extendable beyond *the survey context*. If the moral satisfaction primarily occurs when responding to the survey questions, those who do not belong to the sample (i.e. the majority) would not obtain this improvement in welfare.

a small environmental improvement can be written as:

$$MWTP_E^A = \frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^A}{\partial x^A}} \tag{2}$$

From this expression we have:

RESULT 1. *If Alice and Bob respond honestly to a referendum-type CV question, and if both of them have preferences according to (1), then their MWTPs would be exactly the same as if they had no pure altruistic concerns. However, both the presence of paternalistic and that of animal-focused altruism, increase their MWTP.*

Hence, the *MWTP* expression looks exactly the same as it would have looked in the absence of any pure, i.e. utility focused, altruism between the individuals. Several papers contain similar results using slightly different models (e.g. Bergstrom, 2006; McConnell, 1997). However, first note that the paternalistic environment-focused altruism does not disappear. The reason, of course, is the fact that others payments for *E* do not then offset this type of altruistic utility (which is independent on the other individual's income). Similarly, the non-paternalistic altruism towards animals does not in any way disappear either, since the animals are not expected to pay any off-setting charge corresponding to the increase in *E*.

In order to be able to say anything about whether the conventional cost-benefit criterion will change or not, we must of course also analyse the social optimality condition. In this section we assume that the objective for the government is to maximise a standard Bergson-Samuelson social welfare function (SWF), which can be written as

$$W = w(U^A, U^B), \tag{3}$$

where *w* is increasing in its arguments (i.e. is Paretian) and weakly quasi-concave. An SWF

that can be written as (3) is said to be both welfaristic and anthropocentric, i.e. it depends solely on utility information (welfarism) and it is concerned exclusively with human utility (anthropocentrism). It can be shown (see appendix) that a Pareto efficient allocation then implies that:

$$\frac{\frac{\partial u^A}{\partial E}+\frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}}+\frac{\frac{\partial u^B}{\partial E}+\frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}}=P_E \qquad (4)$$

Combining (4) with (2), and the corresponding *MWTP* expression for Bob, then implies:

$$MWTP_E^A + MWTP_E^B = P_E \qquad (5)$$

which is the basic Samuelson (1954) rule in terms of people's *MWTP*s for the environmental good *E*. Hence, the sum of people's marginal willingness to pay for the good equals the per unit price of it. Note that for this result we do not need to assume anything about *w* beyond the Paretian property. Hence, *w* can here be seen as ordinal, and we need not assume anything with respect to interpersonal comparisons of utility. Nor do we need to assume anything of the utility function (1) beyond ordinality. Thus, we have:

RESULT 2. *If Alice and Bob respond honestly to referendum-type CV questions, and if they have preferences according to (1), then the socially efficient decision rule is to follow the basic Samuelson rule in terms of MWTPs, irrespective of the magnitudes of pure, paternalistic and animal-focused altruism, respectively.*

Note that the fact that Alice's and Bob's *MWTP* increases due to between-people paternalistic altruism as well animal-directed non-paternalistic altruism does not have any implication of the social decision rule. The reason is that this is off-set by the corresponding changes of the Pareto efficiency conditions.

Consequently, the conventional efficiency rule is also the same as it would have been without any between-person pure altruism, implying that if people behave as is assumed here, we can rely on the basic cost-benefit rule to obtain efficiency. Similar results in slightly different contexts, without for example considering animal well-being, are derived by Bergstrom (1999, 2006) and Jones-Lee (1992); see also McConnell (1997) or Johansson-Stenman (1998) in the environmental valuation literature.

3. The Standard Social Welfare Model when Alice and Bob Act as Citizens

One may question whether people really respond to *WTP* questions as utility maximising consumers. A respondent might, for example, focus instead on his individual SWF and respond more as a responsible citizen than as a utility-maximising consumer. This has been suggested by several authors, including philosophers Elizabeth Anderson (1992) and Mark Sagoff (1988, 2004).[8] Although this possibility has been discussed in parts of the economics literature, as far as the author knows, it has not been analysed formally.

Of course, for such a claim to make sense, one must distinguish between utility and choice, i.e. utility cannot simply be defined as what is implicitly maximised by one's choices (in which case people would be utility-maximisers by definition). The standard economic model since Samuelson (1938) does not make such a distinction. As remarked by Broome (1999, p. 4): "Welfare economists move, almost without noticing it, between saying a person prefers one thing to another and saying she is better off with the first than with the second." Of course, in cases where people's choices do reflect the maximisation of their well-being such a distinction is unnecessary. However, when they do not, the question of whether people

[8] Moreover, as argued e.g. by Sen (1977, 1985) and Sugden (1982, 1984), one cannot rule out that the possibility that people sometimes pursue some other end instead of utility (as a measure of individual well-being) for any type of action. This is also the case when taking into account various kinds of altruism.

maximise their utilities or not, is an empirical one. In this paper we define utility throughout to be a measure of individual well-being. Sometimes it will also be a measure reflecting choices, but in this section it will not.[9]

However, perhaps surprisingly, in the conventional case where the SWF is welfaristic and anthropocentric, whether people respond as consumers or citizens may not make much of a difference. Alice's *MWTP* for *E* would then still be given by eq. (2); see appendix for proof. Intuitively, even though Alice now takes Bob's utility completely into account,[10] she would still realise that it would not be in his interest to provide more of *E* than could be justified by motivations of efficiency since additional amounts of *E* would imply higher costs not only to Alice but also to Bob.

RESULT 3. *If Alice and Bob respond truthfully to referendum-type CV questions, and if Alice responds as a citizen and maximises her perceived SWF instead of her utility function, and if Alice and Bob have preferences according to (1), then Alice's MWTP would be exactly the same as if she were responding as a utility-maximising consumer. This is true irrespective of whether Bob responds as a consumer or as a citizen.*

[9] I have argued elsewhere that it is reasonable for the government to be concerned with well-being rather than choices *per se* (Johansson-Stenman 2002), partly influenced by Broome (1991), Harsanyi (1982, 1995) and Ng (1999), which also largely parallels the distinction between decision utility and choice utility discussed by Kahneman et al. (1997). This is also consistent with recent models that allow for paternalism (e.g. Gruber and Köszegi, 2002, 2004; O'Donoghue and Rabin 2003; Thaler and Sunstein, 2003). Nevertheless, it is of course possible to argue and assume that individual choices should be given moral significance *per se*, i.e. independent of individual well-being; see e.g. Arrow (1995) and Sen (1992), and more recently Sugden (2004) for a more systematic analysis.

[10] For example, if *w* is utilitarian then Bob's utility carries the same weight as her own for her stated *MWTP*.

Moreover, the condition for Pareto efficiency is of course unaffected by how people respond to the CV questions. Since, from Result 3, Alice's and Bob's (by symmetry) *MWTP*s are unaffected too, the overall choice rule will also be unaffected. Thus we have:

RESULT 4. *If Alice and Bob respond honestly to referendum-type CV questions, and if they have preferences according to (1), then the socially efficient decision rule is to follow the basic Samuelson rule in terms of MWTPs, irrespective of the magnitudes of pure, paternalistic and animal-focused altruism and irrespective of whether they react to the CV question as social-welfare-maximising citizens or as utility-maximising consumers.*

In other words, the fact that people may act as implicit social planners, or citizens, rather than as consumers does not make any difference in this case. This is equally true for the Pareto efficiency condition, the *MWTP* expressions and the appropriate social decision rule. Consequently, there is no reason to expect either double-counting or that individual *MWTP*s would understate the overall social value. Moreover, we still do not need any information beyond the Paretian property of both an overall SWF and subjectively perceived SWFs, or ordinal properties of the individual utility functions.

4. Modelling Social Welfare beyond Anthropocentrism and Welfarism

So far we have assumed that the underlying social objective is to maximise a conventional welfaristic and anthropocentric SWF. Here we will broaden this objective so that it also values animal welfare and the environment directly. Consider instead the following SWF that is neither anthropocentric nor welfaristic:

$$W = w(U^A, U^B, A, E) \tag{6}$$

where $\frac{\partial w}{\partial U^A} > 0, \frac{\partial w}{\partial U^B} > 0, \frac{\partial w}{\partial A} > 0, \frac{\partial A}{\partial E} > 0$ and where w is quasi-concave. Here animal

suffering clearly counts *per se*, irrespective of whether any human being suffers from observing or knowing about animal suffering.[11] Now, if the overall objective of the government is to maximise (6), this implies that exclusively focusing on obtaining a Pareto efficient allocation ignores the social welfare benefits that are direct, without going through individual human utilities. Formally, the first-best[12] social optimum condition for provision of E is given by (see appendix for proof):

$$\frac{\frac{\partial u^A}{\partial E}+\frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}}+\frac{\frac{\partial u^B}{\partial E}+\frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}}+\Psi\left(\frac{\partial w}{\partial A}\frac{dA}{dE}+\frac{\partial w}{\partial E}\right)=p^E \qquad (7)$$

where

$$\Psi=\frac{1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}}{\frac{\partial w}{\partial U^A}\frac{\partial u^A}{\partial x^A}+\frac{\partial w}{\partial U^B}\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial x^A}}=\frac{1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}}{\frac{\partial w}{\partial U^B}\frac{\partial u^B}{\partial x^B}+\frac{\partial w}{\partial U^A}\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial x^B}} \qquad (8)$$

Hence, this expression is identical to the one based on the standard anthropocentric model, with the exception of the third term. Note also that pure altruism *decreases* the optimum provision of E here. The intuitive reason is that such altruism implies that the utility components in the SWF carry a greater weight than the non-utility elements.

[11] In the classical utilitarian case where animal well-being counts as much as human well-being, supported e.g. by Singer (1975, 1979), we would have $A=\sum_i v^i$ and in our two-individual economy: $W=U^A+U^B+\sum_i v^i$.

[12] Note that "first-best" does not refer to a Pareto efficient allocation here, since we have a broader social objective, but rather that there are no other constraints in addition to the resource constraint.

When Alice and Bob act as consumers

Combining (2), (7) and (8) we have the optimal social decision rule as follows:

$$MWTP_E^A + MWTP_E^B + \Psi\left(\frac{\partial w}{\partial A}\frac{dA}{dE} + \frac{\partial w}{\partial E}\right) = p^E \qquad (9)$$

Since the third term on the right-hand-side of this expression is positive,[13] we directly have:

RESULT 5. *If Alice and Bob respond honestly to referendum-type CV questions as utility-maximising consumers, and if they have preferences according to (1), and if the social objective is to maximise an SWF given by (7), then the optimal provision of environmental quality E exceeds the level given by the basic Samuelson rule in terms of MWTPs.*

Thus, if Alice and Bob behave as we typically assume them to, i.e. as utility maximising individuals, the conventional cost-benefit rule will underestimate the benefit side.

When Alice and Bob respond as welfare-maximising citizens

As mentioned, however, it is not obvious that people would respond to a referendum type CV question as a utility-maximising consumer, even if we take various kinds of altruistic behaviour into account. From the experimental literature it is found that frequently people do not act in a narrow self-interested way alone, but also that, in general, people are not unconditionally altruistic either. Rather, the most promising explanation for the observed behaviour in many types of experimental games, e.g. public good games, dictator games, ultimatum games and trust games, is *conditional co-operation*. Thus, people would like to

[13] Ψ is positive as long as it is optimal for Alice to keep her money, rather than to giving some of them to Bob (see appendix). In the case where it is indeed optimal for Alice to give some of them to Bob, the expression would equal zero (assuming an interior solution).

contribute to a good social cause, and co-operate in public good games, for example, but only if others are doing the same (Fischbacher et al., 2001; Frey and Meier, 2004; Keser and van Winden, 2000; Rabin, 1993). Let us therefore assume that people answer the CV question in order to maximise an SWF provided that others act in the same way. If so, it can formally be shown (see appendix) that Alice's *MWTP* is given by

$$MWTP_E^A = SMRS_{Ex}^A = -\left.\frac{dx^A}{dE}\right|_W = \frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}} + (1-\lambda)\Psi\left(\frac{\partial w}{\partial A}\frac{dA}{dE} + \frac{\partial w}{\partial E}\right) \quad (10)$$

if Bob's *MWTP* is given by:

$$MWTP_E^B = SMRS_{Ex}^B = -\left.\frac{dx^B}{dE}\right|_W = \frac{\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}} + \lambda\Psi\left(\frac{\partial w}{\partial A}\frac{dA}{dE} + \frac{\partial w}{\partial E}\right) \quad (11)$$

where λ is undetermined. Thus, the more *B* is willing to pay (and actually is paying), the less *A* is willing to pay, and vice versa. The mathematical structure of this seems to be the same as in a non-symmetric zero-sum bargaining game. However, the analogy is not perfect since the objective for Alice here is not to maximise her own individual utility (or something else in her own interest such as profits), but rather to maximise social welfare. Nevertheless, the co-ordination problem remains the same. Assuming that this co-ordination problem can somehow be solved, the sum of Alice's and Bob's *MWTP* can be uniquely determined as follows:

$$MWTP_E^A + MWTP_E^B = \frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}} + \frac{\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}} + \Psi\left(\frac{\partial w}{\partial A}\frac{dA}{dE} + \frac{\partial w}{\partial E}\right) \quad (12)$$

Now, what can be said about the choice rule? Combining (9) and (12) it directly follows that

$MWTP_E^A + MWTP_E^B = p^E$, i.e. the basic Samuelson rule in terms of *MWTP*s.

Remember that this result was derived based on conditional cooperation in the specific sense that Alice responds in order to maximize social welfare conditional on the fact that Bob does the same. One might then want to know what happens if Alice responds *unconditionally* of how Bob's acts. Assume therefore that Bob responds as a utility-maximising consumer and that Alice knows this. In this case it can be shown (see appendix) that (10) and (11) continue to hold for the special case where $\lambda = 0$, which also implies that (12) continues to hold. Consequently, it is not necessary for both of them to respond as citizens, for the Samuelson rule to hold. Thus, we have:

RESULT 6. *If Alice and Bob respond honestly to referendum-type CV questions and at least one of them responds as a citizen, and if the social objective is to maximise an SWF given by (7), then the socially optimal decision rule is to follow the basic Samuelson rule in terms of MWTPs.*

Consequently, if people have non-welfaristic ethical preferences and respond to the WTP question as a welfare-maximising citizens, it may still be optimal to use the basic cost-benefit criteria, conditional on the fact that the government respects the ethical preferences of its citizens. This is not obvious though. It may be that the government maintains its right to rely on its own ethical views that may differ from its citizens' views. If so, the marginal social value may be lower than the sum of the marginal WTP, to the extent that these reflect non-anthropocentric benefits. However, as shown, it is also possible that the government does respect potential non-anthropocentric ethical views of its citizens, but that the respondents to CV-studies do not act as utility-maximising consumers, rather than as citizens, implying that the marginal social value may be higher than the sum of the *MWTP*s.

5. Conclusion

In this paper we have analysed implications of non-anthropocentric and non-welfaristic preferences, as well as cases where people respond to WTP questions as utility-maximising consumers and welfare-maximising citizens, respectively. As far as the author knows, this is the first study that systematically explores theoretical implications of non-welfarism and non-anthropocentrism. Still, as noted in the introduction, the benefits in terms of increased realism do not necessarily outweigh the costs associated with more complicated models. The results here, however, illustrate that these costs are sometimes smaller than one might expect. Indeed, it was shown that in many of the cases analysed, the basic efficiency rule (the Samuelson rule) in terms of aggregate marginal willingness to pay still holds. From a policy-making perspective this simplifies life greatly because, for example, the extent to which people are altruistic in different ways can be ignored. Nor do we have to care about the structure of either the social welfare function or the utility function (beyond monotonicity and quasi concavity). Needless to say, however, this fact should not prevent us from realising that it is sometimes inappropriate to use the basic efficiency rule. However, even though the modified choice rules are in most cases much more difficult to interpret in monetary terms, they do at least provide guidance about whether the basic efficiency rule implies an over or an under provision. Finally, this is a first attempt to incorporate non-anthropocentric and non-welfaristic assumptions into welfare economics, and there is, of course, room for extension in many dimensions.

Appendix

Proof of equation (2)

We first derive an expression for the individual *MWTP*s, and then derive the social optimum condition. Totally differentiating (2) and setting du^1 equal to zero, we get:

$$
\begin{aligned}
du^A &= \frac{\partial u^A}{\partial x^A}dx^A + \left(\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}\right)dE + \frac{\partial u^A}{\partial u^B}du^B \\
&= \frac{\partial u^A}{\partial x^A}dx^A + \left(\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}\right)dE \\
&+ \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B + \left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}\right)dE + \frac{\partial u^B}{\partial u^A}du^A\right) \qquad \text{(A1)} \\
&= \frac{\partial u^A}{\partial x^A}dx^A + \left(\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}\right)dE \\
&+ \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B + \left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}\right)dE\right) = 0
\end{aligned}
$$

where in the last step we have thus used that $du^A = 0$. Then it follows that if

$$dx^B = -\left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}\right)\Big/\frac{\partial u^B}{\partial x^B}dE \text{ then}$$

$$\frac{\partial u^A}{\partial x^A}dx^A + \left(\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}\right)dE = 0 \qquad \text{(A2)}$$

Alice's *MWTP* for *E* is then given by:

$$MWTP_E^A = MRS_{Ex}^A = -\left.\frac{dx^A}{dE}\right|_{u^1} = \frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^A}{\partial x^A}} \qquad \text{(A3)}$$

Then, by symmetry, we of course also have

$$MWTP_E^B = MRS_{Ex}^B = -\left.\frac{dx^B}{dE}\right|_{u^1} = \frac{\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^B}{\partial x^B}} \qquad \text{(A4)}$$

End of proof.

Proof of equation (4)

The Pareto efficient allocation is found by maximising u^A subject to a resource or production constraint, which for simplicity we assume to be linear, while holding u^B constant, i.e. max. $u^A = u^A(x^A, E, A(E), u^B, E^B(E))$ s.t. $u^B = \overline{u^B}$ and the resource constraint. But since u^B is held constant, this is equivalent to the maximisation of $u^A = u^A(x^A, E, A(E), \overline{u^B}, E^B(E))$ subject to the same constraints, implying the Lagrangean:[14]

$$u^A(x^A, E, A(E), \overline{u^B}, E^B(E)) + \lambda(u^B(x^B, E, A(E), \overline{u^A}, E^A(E)) - \overline{u^B}) + \mu(R - x^A - x^B - p_E E) \tag{A5}$$

and the corresponding first order conditions for an interior optimum:

$$\frac{\partial u^A}{\partial x^A} = \lambda \frac{\partial u^B}{\partial x^B} = \mu \tag{A6}$$

$$\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E} + \lambda\left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}\right) = \mu p_E \tag{A7}$$

where (A6) and (A7) directly imply that

$$\frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}} + \frac{\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}} = P_E \tag{A8}$$

End of proof.

[14] Where u^A is held constant in Bob's utility function due to the envelope theorem.

Proof that respondents who act like citizens and maximise social welfare would imply the *MWTP* in eq. (3)

Substituting (1) into (3) for both Alice and Bob implies that:

$$W = w(u^A(x^A, E, A(E), u^B, E^B(E)), u^B(x^B, E, A(E), u^A, E^A(E))) \tag{A9}$$

From totally differentiating (A9) it follows that

$$\begin{aligned}
dW &= \frac{\partial w}{\partial u^A} du^A + \frac{\partial w}{\partial u^B} du^B = \frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE + \frac{\partial u^A}{\partial u^B} du^B\right) + \frac{\partial w}{\partial u^B} du^B \\
&= \frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE + \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B} dx^B + \Omega^B dE + \frac{\partial u^B}{\partial u^A} du^A\right)\right) \\
&+ \frac{\partial w}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B} dx^B + \Omega^B dE + \frac{\partial u^B}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE + \frac{\partial u^A}{\partial u^B} du^B\right)\right) \\
&= \frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE + \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B} dx^B + \Omega^B dE\right)\right) \\
&+ \frac{\partial w}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B} dx^B + \Omega^B dE + \frac{\partial u^B}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE\right)\right) = 0
\end{aligned} \tag{A10}$$

where $\Omega^A = \frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}$ and $\Omega^B = \frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}$ and where in the last step we used that $\frac{\partial w}{\partial u^A} du^A + \frac{\partial w}{\partial u^B} du^B = 0$. Then it follows that if

$$dx^B = -\frac{\Omega^B}{\frac{\partial u^B}{\partial x^B}} dE \tag{A11}$$

then (A10) can be rewritten as:

$$\begin{aligned}
&\frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE\right) + \frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE\right) \\
&= \left(\frac{\partial w}{\partial u^A} + \frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)\left(\frac{\partial u^A}{\partial x^A} dx^A + \Omega^A dE\right) = 0
\end{aligned} \tag{A12}$$

implying that

$$MWTP_E^A = SMRS_{Ex}^A = -\left.\frac{dx^A}{dE}\right|_W = \frac{\Omega^A}{\frac{\partial u^A}{\partial x^A}} = \frac{\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}} \tag{A13}$$

where $SMRS_{Ex}^A$ is Alice's perceived social marginal rate of substitution between E and x^A. This expression is clearly identical to eq. (3). Note that this expression holds if (and only if) (A11) holds, i.e. if

$$MWTP_E^B = SMRS_{Ex}^B = -\left.\frac{dx^B}{dE}\right|_W = \frac{\Omega^B}{\frac{\partial u^B}{\partial x^B}} = \frac{\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}} \tag{A14}$$

Since (A13) and (A14) hold simultaneously, conditional on each other, they represent a Nash equilibrium. End of proof.

Proof of eq. (7)

Maximising social welfare conditional on the budget constraint implies the Lagrangean

$$w(U^A, U^B, A, E) + \mu(R - x^A - x^B - p_E E) \tag{A15}$$

and the corresponding first order conditions for an interior solution:

$$\frac{dw}{dx^A} - \mu = 0 \tag{A16}$$

$$\frac{dw}{dx^B} - \mu = 0 \tag{A17}$$

$$\frac{dw}{dE} - p_E \mu = 0 \tag{A18}$$

where for example $\frac{dw}{dx^A}$ is the total welfare effect per unit of increased x^A, i.e. including all different mechanisms (such as Alice's direct effect, Bob's increased utility from Alice's increased utility, Alice's increased utility from Bob's increased utility due to Alice's increased utility etc.), holding consumption of other goods constant. Then we have

$$\frac{dw}{dx^A} = \frac{\partial w}{\partial U^A}\frac{du^A}{dx^A} + \frac{\partial w}{\partial U^B}\frac{du^B}{dx^A} \tag{A19}$$

where similarly $\frac{du^A}{dx^A}$ is the total utility increase for Alice per unit of increased x^A, including all different mechanisms and holding consumption of other goods constant. Thus we have

$$\frac{du^A}{dx^A} = \frac{\partial u^A}{\partial x^A} + \frac{\partial u^A}{\partial u^B}\frac{du^B}{dx^A} \tag{A20}$$

and correspondingly

$$\frac{du^B}{dx^A} = \frac{\partial u^B}{\partial u^A}\frac{du^A}{dx^A} \tag{A21}$$

Substituting (A21) into (A20) gives

$$\frac{du^A}{dx^A} = \frac{\partial u^A}{\partial x^A} + \frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\frac{du^A}{dx^A} = \frac{\partial u^A}{\partial x^A}\Bigg/\left(1 - \frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right) = \alpha\frac{\partial u^A}{\partial x^A} \tag{A22}$$

where $\alpha = \left(1 - \frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)^{-1}$

Substituting (A22) into (A21) yields:

$$\frac{du^B}{dx^A} = \alpha\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial x^A} \tag{A23}$$

Substituting (A22) and (A23) into (A21) then implies that

$$\frac{dw}{dx^A} = \alpha\left(\frac{\partial w}{\partial U^A}\frac{\partial u^A}{\partial x^A} + \frac{\partial w}{\partial U^B}\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial x^A}\right) \tag{A24}$$

By symmetry we also have

$$\frac{dw}{dx^B} = \alpha\left(\frac{\partial w}{\partial U^A}\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial x^B} + \frac{\partial w}{\partial U^B}\frac{\partial u^B}{\partial x^B}\right) \tag{A25}$$

We also have

$$\frac{dw}{dE} = \frac{\partial w}{\partial U^A}\frac{du^A}{dE} + \frac{\partial w}{\partial U^B}\frac{du^B}{dE} + \frac{\partial w}{\partial A}\frac{dE}{dE} + \frac{\partial w}{\partial E} \tag{A26}$$

where

$$\frac{du^A}{dE} = \frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial u^B}\frac{du^B}{dE} + \frac{\partial u^A}{\partial E^B}\frac{dE^B}{dE} \tag{A27}$$

and

$$\frac{du^B}{dE} = \frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial u^A}\frac{du^A}{dE} + \frac{\partial u^B}{\partial E^A}\frac{dE^A}{dE} \tag{A28}$$

Substituting (A28) into (A27) implies:

$$\begin{aligned}
&\frac{du^A}{dE} = \frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{dE^B}{dE} \\
&+ \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial u^A}\frac{du^A}{dE} + \frac{\partial u^B}{\partial E^A}\frac{dE^A}{dE}\right) \\
&= \alpha\left(\frac{\partial u^A}{\partial E} + \frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^A}{\partial E^B}\frac{dE^B}{dE} + \frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial E} + \frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E} + \frac{\partial u^B}{\partial E^A}\frac{dE^A}{dE}\right)\right) \\
&= \alpha\left(\Omega^A + \frac{\partial u^A}{\partial u^B}\Omega^B\right)
\end{aligned} \tag{A29}$$

Hence, by symmetry

$$\frac{du^B}{dE} = \alpha\left(\Omega^B + \frac{\partial u^B}{\partial u^A}\Omega^A\right) \tag{A30}$$

Thus, from combining (A26), (A29) and (A30) we have

$$\frac{dw}{dE} = \frac{\partial w}{\partial A}\frac{dA}{dE} + \frac{\partial w}{\partial E} + \frac{\partial w}{\partial U^A}\alpha\left(\Omega^A + \frac{\partial u^A}{\partial u^B}\Omega^B\right) + \frac{\partial w}{\partial U^B}\alpha\left(\Omega^B + \frac{\partial u^B}{\partial u^A}\Omega^A\right) \tag{A31}$$

Combining (A16), (A17) and (A18) we have

$$\frac{\dfrac{dw}{dE}}{\dfrac{dw}{dx^B}} = \frac{\dfrac{dw}{dE}}{\dfrac{dw}{dx^A}} = p^E \tag{A32}$$

Substituting (A24) and (A31) into (A32) and some standard algebraical manipulations implies:

$$\frac{\dfrac{dw}{dE}}{\dfrac{dw}{dx^A}} = \frac{\Omega^A}{\dfrac{\partial u^A}{\partial x^A}} + \frac{\Omega^B}{\dfrac{\partial u^B}{\partial x^B}} + \frac{\dfrac{\partial w}{\partial A}\dfrac{dA}{dE} + \dfrac{\partial w}{\partial E}}{\dfrac{\partial w}{\partial U^A}\dfrac{\partial u^A}{\partial x^A} + \dfrac{\partial w}{\partial U^B}\dfrac{\partial u^B}{\partial u^A}\dfrac{\partial u^A}{\partial x^A}}\left(1 - \frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right) = p^E \tag{A33}$$

Finally, imposing the expressions for Ω^A and Ω^B yields:

$$\frac{\frac{\partial u^A}{\partial E}+\frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}}+\frac{\frac{\partial u^B}{\partial E}+\frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}}$$
$$+\frac{\frac{\partial w}{\partial A}\frac{dA}{dE}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial U^A}\frac{\partial u^A}{\partial x^A}+\frac{\partial w}{\partial U^B}\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial x^A}}\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)=p^E \tag{A34}$$

End of proof.

<u>Proof of eqs. (10) and (11)</u>

Totally differentiating (A9) implies

$$\begin{aligned}dW&=\frac{\partial w}{\partial u^A}du^A+\frac{\partial w}{\partial u^B}du^B+\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE+\frac{\partial w}{\partial E}dE\\&=\frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A}dx^A+\Omega^A dE+\frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B+\Omega^B dE+\frac{\partial u^B}{\partial u^A}du^A\right)\right)\\&+\frac{\partial w}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B+\Omega^B dE+\frac{\partial u^B}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A}dx^A+\Omega^A dE+\frac{\partial u^A}{\partial u^B}du^B\right)\right)\\&+\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE+\frac{\partial w}{\partial E}dE=0\end{aligned} \tag{A35}$$

Using that $\frac{\partial w}{\partial u^A}du^A+\frac{\partial w}{\partial u^B}du^B=-\left(\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE+\frac{\partial w}{\partial E}dE\right)$ we can rewrite (A35) as

$$\begin{aligned}dW&=\frac{\partial w}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A}dx^A+\Omega^A dE+\frac{\partial u^A}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B+\Omega^B dE\right)\right)\\&+\frac{\partial w}{\partial u^B}\left(\frac{\partial u^B}{\partial x^B}dx^B+\Omega^B dE+\frac{\partial u^B}{\partial u^A}\left(\frac{\partial u^A}{\partial x^A}dx^A+\Omega^A dE\right)\right)\\&+\left(\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE+\frac{\partial w}{\partial E}dE\right)\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)=0\end{aligned} \tag{A36}$$

Then it follows that if

$$dx^B = -\frac{\Omega^B}{\frac{\partial u^B}{\partial x^B}}dE - \lambda\frac{1}{\frac{\partial u^B}{\partial x^B}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^A}\frac{\partial u^A}{\partial u^B}+\frac{\partial w}{\partial u^B}}\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)dE \tag{A37}$$

then (A36) can be rewritten as

$$\begin{aligned} dW &= \left(\frac{\partial w}{\partial u^A}+\frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)\left(\frac{\partial u^A}{\partial x^A}dx^A+\Omega^A dE\right) \\ &+(1-\lambda)\left(\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE+\frac{\partial w}{\partial E}dE\right)\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)=0 \end{aligned} \tag{A38}$$

so that

$$dx^A = -\frac{\Omega^A}{\frac{\partial u^A}{\partial x^A}}dE-(1-\lambda)\frac{1}{\frac{\partial u^A}{\partial x^A}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}+\frac{\partial w}{\partial u^A}}\left(1-\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right)dE \tag{A39}$$

This implies by definition that

$$\begin{aligned} MWTP_E^A = SMRS_{Ex}^A &= -\left.\frac{dx^A}{dE}\right|_W = \frac{\Omega^A}{\frac{\partial u^A}{\partial x^A}}+(1-\lambda)\frac{1}{\frac{\partial u^A}{\partial x^A}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}+\frac{\partial w}{\partial u^A}}\left(1-\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right) \\ &= \frac{\frac{\partial u^A}{\partial E}+\frac{\partial u^A}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^A}{\partial E^B}\frac{\partial E^B}{\partial E}}{\frac{\partial u^A}{\partial x^A}}+(1-\lambda)\frac{1}{\frac{\partial u^A}{\partial x^A}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}+\frac{\partial w}{\partial u^A}}\left(1-\frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right) \end{aligned} \tag{A40}$$

Note that this holds conditional on (A37), which in turn implies that

$$\begin{aligned} MWTP_E^B = SMRS_{Ex}^B &= -\left.\frac{dx^B}{dE}\right|_W = \frac{\Omega^B}{\frac{\partial u^B}{\partial x^B}}+\lambda\frac{1}{\frac{\partial u^B}{\partial x^B}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^A}\frac{\partial u^A}{\partial u^B}+\frac{\partial w}{\partial u^B}}\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right) \\ &= \frac{\frac{\partial u^B}{\partial E}+\frac{\partial u^B}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial u^B}{\partial E^A}\frac{\partial E^A}{\partial E}}{\frac{\partial u^B}{\partial x^B}}+\lambda\frac{1}{\frac{\partial u^B}{\partial x^B}}\frac{\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}+\frac{\partial w}{\partial E}}{\frac{\partial w}{\partial u^A}\frac{\partial u^A}{\partial u^B}+\frac{\partial w}{\partial u^B}}\left(1-\frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right) \end{aligned} \tag{A41}$$

Since (A40) and (A41) hold simultaneously, conditional on each other, they represent a Nash equilibrium. End of proof.

Proof of eqs. (10) and (11) when Bob respond as a utility-maximising consumer

A(36) holds as above. If

$$dx^B = -\frac{\Omega^B}{\dfrac{\partial u^B}{\partial x^B}}dE \tag{A42}$$

then (A36) can be rewritten as

$$\begin{aligned} dW &= \left(\frac{\partial w}{\partial u^A} + \frac{\partial w}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right)\left(\frac{\partial u^A}{\partial x^A}dx^A + \Omega^A dE\right) \\ &+ \left(\frac{\partial w}{\partial A}\frac{\partial A}{\partial E}dE + \frac{\partial w}{\partial E}dE\right)\left(1 - \frac{\partial u^A}{\partial u^B}\frac{\partial u^B}{\partial u^A}\right) = 0 \end{aligned} \tag{A43}$$

so that

$$dx^A = -\frac{\Omega^A}{\dfrac{\partial u^A}{\partial x^A}}dE - \frac{1}{\dfrac{\partial u^A}{\partial x^A}}\frac{\dfrac{\partial w}{\partial A}\dfrac{\partial A}{\partial E} + \dfrac{\partial w}{\partial E}}{\dfrac{\partial w}{\partial u^B}\dfrac{\partial u^B}{\partial u^A} + \dfrac{\partial w}{\partial u^A}}\left(1 - \frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right)dE \tag{A44}$$

and

$$\begin{aligned} MWTP_E^A = SMRS_{Ex}^A &= -\left.\frac{dx^A}{dE}\right|_W = \frac{\Omega^A}{\dfrac{\partial u^A}{\partial x^A}} + \frac{1}{\dfrac{\partial u^A}{\partial x^A}}\frac{\dfrac{\partial w}{\partial A}\dfrac{\partial A}{\partial E} + \dfrac{\partial w}{\partial E}}{\dfrac{\partial w}{\partial u^B}\dfrac{\partial u^B}{\partial u^A} + \dfrac{\partial w}{\partial u^A}}\left(1 - \frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right) \\ &= \frac{\dfrac{\partial u^A}{\partial E} + \dfrac{\partial u^A}{\partial A}\dfrac{\partial A}{\partial E} + \dfrac{\partial u^A}{\partial E^B}\dfrac{\partial E^B}{\partial E}}{\dfrac{\partial u^A}{\partial x^A}} + \frac{1}{\dfrac{\partial u^A}{\partial x^A}}\frac{\dfrac{\partial w}{\partial A}\dfrac{\partial A}{\partial E} + \dfrac{\partial w}{\partial E}}{\dfrac{\partial w}{\partial u^B}\dfrac{\partial u^B}{\partial u^A} + \dfrac{\partial w}{\partial u^A}}\left(1 - \frac{\partial u^B}{\partial u^A}\frac{\partial u^A}{\partial u^B}\right) \end{aligned} \tag{A45}$$

This holds conditional on (A42), which in turn implies that

$$MWTP_E^B = SMRS_{Ex}^B = -\left.\frac{dx^B}{dE}\right|_W = \frac{\Omega^B}{\dfrac{\partial u^B}{\partial x^B}} = \frac{\dfrac{\partial u^B}{\partial E} + \dfrac{\partial u^B}{\partial A}\dfrac{\partial A}{\partial E} + \dfrac{\partial u^B}{\partial E^A}\dfrac{\partial E^A}{\partial E}}{\dfrac{\partial u^B}{\partial x^B}} \tag{A46}$$

Since (A45) and (A46) hold simultaneously, conditional on each other, they represent a Nash equilibrium. End of proof.

References

Anderson, E. (1993) *Value in Ethics and in Economics*. Cambridge, Mass. Harvard University Press.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence, *Journal of Political Economy* 97(6), 1447-58.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, *Economic Journal*, 100, 464-477.

Archibald, G. C. and Donaldson, D. (1976), 'Non-paternalism and the Basic Theorems of Welfare Economics', *Canadian Journal of Economics* **9**, 492-507.

Arrow, Kenneth J. (1995). 'A note on freedom and flexibility'. In Kaushik Basu, Prasanta Pattanaik and Kotaro Suzumura (eds), *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen*, pp. 7-16. Oxford: Oxford University Press, pp. 7-16.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman (1993), *Report of the NOAA Panel on Contingent Valuation*, Resources For the Future, Washington, D.C.

Bergstrom, T.C. (1999). Systems of benevolent utility functions. *Journal of Public Economic Theory* 1, 71-100.

Bergstrom, T.C. (2006). Benefit-Cost in a Benevolent Society. *American Economic Review*, 96, 339-51.

Boadway, R. and M. Keen. 1993. Public Goods, Self-Selection and Optimal Income Taxation. *International Economic Review* 34: 463-78.

Broome, J. (1991), *Weighing Goods*, Oxford: Blackwell.

Broome, J. (1999), *Ethics out of Economics*, Cambridge: Cambridge University Press.

Diamond, P. and Hausman, J. (1994). Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives*, 8: 45-64.

Fischbacher U., S. Gächter and E. Fehr (2001). Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, *Economic Letters* 71, 397- 404.

Frey, B. S. and S. Meier (2004). Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment, *American Economic Review*, 94(5), 1717-22.

Gruber, J. and B. Köszegi (2002) Is addiction "rational"? Theory and evidence, *Quarterly Journal of Economics*, 116(4), 1261-1303.

Gruber, J. and B. Köszegi (2004), Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes, *Journal of Public Economics*, 88(9-10), 1959-87.

Harrison, G. W. (1992). Valuing Public Goods with the Contingent Valuation Method: A Critique. *Journal of Environmental Economics and Management* 23, 248-257.

Harsanyi, J.C. (1982), 'Morality and the theory of rational behavior', in Sen and Williams (1982).

Harsanyi, J.C. (1995), 'A Theory of Prudential Values and a Rule Utilitarian Theory of Morality', *Social Choice and Welfare*, 12(4), 319-33.

Jacobsson, F., M. Johannesson and L. Borgqvist (2006) Is altruism paternalistic? *Economic Journal*, forthcoming.

Johansson-Stenman, O. (1998), 'The importance of ethics in environmental economics with a focus on existence values', *Environmental and Resource Economics*, 11(3-4), 429-42.

Johansson-Stenman, O. (2002), What should we do with inconsistent, non-welfaristic and undeveloped preferences? in Bromley and Paavola (eds.) *Economics, Ethics, and Environmental Policy: Contested Choices*, Blackwell.

Johansson-Stenman, O. (2005) "Global environmental problems, efficiency and limited altruism" *Economics Letters*, 86(1), 101-106.

Johansson-Stenman, O. (2006), Should Animal Welfare Count?, working paper 197,

Department of Economics, Göteborg University.

Jones-Lee, M.W. (1992), 'Paternalistic Altruism and the Value of a Statistical Life', *Economic Journal* 102, 80-90.

Kahneman, D. and J. L. Knetsch (1992), 'Valuing public goods: the purchase of moral satisfaction', *Journal of Environmental Economics and Management,* 22(1), 57-70.

Kahneman, D., P. P. Wakker and R. Sarin (1997) Back to Bentham? Explorations of Experienced Utility, *Quarterly Journal of Economics*, 112(2), 375-405.

Kahneman, D., Ritov, I. and Schkade, D. (1999). 'Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues', *Journal of Risk and Uncertainty*, 19, 203-235.

Kaplow, L. and S. Shavell (2001), Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle. *Journal of Political Economy,* 109, 281-286.

Keser, C. and F van Winden (2000) Conditional Cooperation and Voluntary Contributions to Public Goods, *Scandinavian Journal of Economics*, 102(1), 23-39.

Marshall, A.S. (1890) *Principles of Economics*. London: Macmillan.

McConnell, K.E. (1997), 'Does Altruism Undermine Existence Value?', *Journal of Environmental Economics and Management* **32**(1), 22-37.

Ng, Y.-K. (1999), 'Utility, Informed Preference, or Happiness: Following Harsanyi's Argument to Its Logical Conclusion', *Social Choice and Welfare*, 16(2), 197-216.

O'Donoghue, T. and M. Rabin (2003) Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes *American Economic Review, papers and proceedings,* 93(2), 186-91.

Rabin, M. (1993), 'Incorporating Fairness into Game Theory and Economics', *American Economic Review* **83**(5), 1281-1302.

Sagoff, M. (1988) *The Economy of the Earth: Philosophy, Law, and the Environment, Cambridge: Cambridge University Press.*

Sagoff, M. (2004) *Price, Principle, and the Environment, Cambridge: Cambridge University Press.*

Samuelson, P. A. (1938), 'A note on the pure theory of consumer behaviour', *Econometrica* 5, 61-71.

Samuelson, P.A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics* 36, 387-389.

Sen, A. K. (1970). *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Sen, A. K. (1977) Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs*, 6, 317-344.

Sen, A. K. (1979), 'Personal utilities and public judgements: or what's wrong with welfare economics?' *Economic Journal*, 89, 537-58.

Sen, A. K. (1985a). *Commodities and capabilities*. Amsterdam: North-Holland.

Sen, A. K. (1985b), 'Goals, Commitment, and Identity', *Journal of Law, Economics and Organization*, 1(2), 341-55.

Sen, A. K. (1987), *On Ethics & Economics*, Blackwell.

Sen, A.K. (1992), *Inequality Reexamined*, Cambridge Ma.: Harvard University Press.

Sen, A.K. (1993). Capability and well-being. In M. Nussbaum & A.K. Sen (Eds), *The Quality of life*. Oxford, Clarendon Paperbacks. Oxford University Press

Singer, P. (1975), *Animal Liberation: A New Ethics for Our Treatment of Animals*, New York: Avon.

Singer, P. (1979), *Practical Ethics*, Cambridge University Press.

Sugden, R. (2004), The Opportunity Criterion: Consumer Sovereignty Without the assumption of Coherent Preferences *American Economic Review*, 94(4), 1014-33.

Thaler, R. H. and C. R. Sunstein (2003), Libertarian Paternalism, *American Economic Review*, papers and proceedings, 93(2), 175-79.