

Morse-Jones, Sian; Bateman, Ian; Kontoleon, Andreas; Ferrini, Silvia; Burgess, Neil

Working Paper

Testing the theoretical consistency of stated preferences for tropical wildlife conservation

CSERGE Working Paper EDM, No. 10-02

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Morse-Jones, Sian; Bateman, Ian; Kontoleon, Andreas; Ferrini, Silvia; Burgess, Neil (2010) : Testing the theoretical consistency of stated preferences for tropical wildlife conservation, CSERGE Working Paper EDM, No. 10-02, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/48814>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Testing the theoretical consistency of stated preferences for tropical wildlife conservation

by

Sian Morse-Jones, Ian Bateman, Andreas Kontoleon,
Silvia Ferrini, Neil Burgess and Kerry Turner

CSERGE Working Paper EDM 10-02

**TESTING THE THEORETICAL CONSISTENCY OF STATED PREFERENCES
FOR TROPICAL WILDLIFE CONSERVATION**

**Sian Morse-Jones¹, Ian Bateman¹, Andreas Kontoleon², Silvia Ferrini¹, Neil Burgess³
and Kerry Turner¹**

¹CSERGE, University of East Anglia, Norwich

²Department of Land Economy, University of Cambridge

³Department of Zoology, University of Cambridge

Corresponding author: Sian Morse-Jones: s.morse-jones@uea.ac.uk

CSERGE working paper

Environmental Decision Making (EDM) Series: Environmental Valuation

ISSN 0967-8875

Abstract

In this paper, we report the results of an online choice experiment designed to test the theoretical consistency of stated preferences for a complex predominantly non-use good. Our case study concerns the values held by UK residents for the conservation of wildlife in the Eastern Arc Mountains in Tanzania, part of the Eastern Afromontane “biodiversity hotspot”. Theoretical consistency is assessed through tests of value sensitivity to the scope of the scheme, ordering effects and (more unusually) the presence of substitutes. Critically, we find that ‘substitution effects’ may account for apparent insensitivity to scope. We also find some evidence that preferences are invariant to the presentation order. Our results highlight the importance of incorporating information about substitutes into the valuation process since welfare estimates may otherwise be overstated.

Key words: environmental valuation; choice experiment; scope tests; substitution effects; order effects.

Acknowledgements: We are grateful to Brendan Fisher, Robin Naidoo and Carlo Fezzi for their valuable input on earlier stages of this work. We would also like to thank Michele Menegon and David C. Blackburn for kindly allowing the use of their images in the survey instrument. This work is part of the Valuing the Arc programme (www.valuingthearc.org) and was supported by the Leverhulme Trust.

1 INTRODUCTION

Tropical biodiversity is declining at unprecedented rates (CBD, 2006) and urgent action is required to safeguard against further losses. However, conservation efforts in many developing countries are severely constrained by large shortfalls in available funding (Balmford et al, 2003). Correspondingly, policymakers have called for a better understanding of the value attached to tropical wildlife conservation (CBD, 2005).

To date, very few empirical studies have been conducted to estimate the benefits of conserving tropical biodiversity to residents in 'donor' countries (see, for example, Horton et al 2003; Kramer and Mercer, 1997; Rolfe et al 2000). Obtaining valid and reliable values for such distant and complex goods raises a number of methodological challenges to valuation practitioners and presents a more cognitively demanding task to respondents than is encountered under the elicitation of values for local, familiar and/or less complex goods. An initial concern is whether, when faced with previously unknown and complex goods respondents are able to 'discover' theoretically consistent preferences (as per Plott, 1996) or instead use available heuristics to 'construct' responses which are susceptible to framing effects and thereby fail tests of procedural invariance (as per Slovic, 1995).

In seeking to distinguish between the former theoretically consistent preferences and those that are inadmissible within wider cost benefit analyses (CBA) the conventional 'gold standard' test is to examine the scope sensitivity of valuations (Arrow et al., 1993). While we administer such a test, we argue that it is flawed and insufficient in that a finding of mere statistical significance of differences as the scope of a good changes does not imply that the degree of sensitivity is reasonable and sufficient. To address this failing we extend our survey design to allow tests for substitution effects. Economic theory provides the clear expectation that the closer the similarity between two substitutes the greater should be the resultant substitution effect. Despite the fundamental nature of such a relation, substitution tests are rarely considered within stated preference (SP) methods and in particular in choice experiment analyses.

In the context of valuing complex and unfamiliar goods such as tropical rainforests, Carson (1998) specifically notes the importance of considering how the provision of substitute conservation programmes might affect the values attached to a specific site. This echoes concerns reported elsewhere in the CVM literature that if respondents are not reminded about other similar goods they may overestimate their WTP for a specific good or instead state the value they hold for the general type of good (Arrow et al. 1993; Loomis et al, 1994). The NOAA panel recommends the use of reminder statements to ensure that respondents give full consideration to substitute goods (Arrow et al. 1993). Yet, these may not always be effective (see, for example, Loomis et al 1994; Ojea & Loureiro 2009), especially where respondents are unfamiliar with the good in question. Approaches which seek to stimulate more explicit consideration of substitutes appear to have been more successful (see Hailu et al, 2000; Neill 1995). In one of the very few empirical studies to elicit preferences for rainforest conservation, Rolfe et al (2000) use a labelled choice experiment to disguise the location of interest amongst a pool of substitutes. Their work provides valuable information with regard to how preferences for a particular conservation site are affected by varying the availability of domestic and overseas sites. However, the authors obtain only *relative* measures of WTP due to the omission of information about the baseline conditions. We build on this work by designing a choice experiment which enables the good of interest to be presented alongside a potential substitute, with full disclosure of baseline conditions for both sites.

We approach the above issues using an online choice experiment survey which elicits marginal willingness to pay (WTP) for choice attributes and absolute welfare measures for

conserving wildlife in the Eastern Arc Mountains either *separately* or as part of a wider '*bundled*' good that includes a potential substitute in the form of wildlife conservation in the Cameroon Highlands. In doing so, we use a *novel design* which enables us to test explicitly whether respondents are externally sensitive to scope and to examine how this is impacted by possible substitution effects. An ad hoc combination of experimental treatments allows us to test the effect of framing the good separately or alongside either similar or dissimilar substitutes. Further treatments allow us also to examine ordering effects. All of the above effects are investigated with respect to preferences for conserving endemic versus non-endemic species, and charismatic versus non-charismatic species.

2 THE SCOPE TEST: LIMITATIONS

A problem identified in the stated preference valuation literature is that very few studies undertake rigorous testing of the theoretical consistency of values. Instead, many rely solely on the use of scope tests. The scope test involves examining whether WTP changes in accordance with changes in the quantity or quality of a good, as economic theory tells us that individuals should be willing to pay more for a larger and/or more inclusive good compared with a component part. The widespread interest in examining scope sensitivity stems from the concern that WTP may be invariant to changes in the quantity or quality of the non-market good and vulnerable to 'embedding effects'. There is mixed evidence of this in the empirical literature with some studies reporting scope sensitivity (see for example, Carson and Mitchell, 1993; Loomis & White, 1996); while others find values to be insensitive to scope (see for example, Boyle et al, 1994; White et al, 1997), and still more find evidence of both sensitivity and insensitivity (see, for example, Giraud et al 1999; Veisten et al 2004 and Horton et al 2001; 2003).

However, there are a number of reasons why many of these scope tests may not provide a reliable indication of the theoretical consistency of preferences (Banerjee & Murphy, 2005). A primary problem with the standard scope test is that it merely tests whether WTP is larger for more of a good, that is, there are only weak expectations with regard to the degree of sensitivity (Fisher, 1996). In essence as long as the good is not a 'bad', then any non-negative answer would pass the test. In this paper we question this logic and ask whether this really provides a good indication of theoretical validity.

While the literature tells us that respondents do adjust their WTP for more (or less) of a good, the problem that remains is that they still may not really comprehend what the good is just that they have been offered more of it. For example, Ariely et al (2003) find evidence of "coherent arbitrariness" even for ordinary consumer products whereby the initial reported WTP values appear to be arbitrarily set (being unduly influenced by non-normative factors including the respondent's social security number), but subsequent WTP responses to larger or smaller versions of the good are internally consistent. Critically, in such situations a finding of sensitivity to scope could falsely support a conclusion of consistent and stable preferences.

Alongside this, there are differing views over how to interpret scope insensitivity. Kahneman and Knetsch (1992) argue that embedding provides evidence that people are not valuing the intended good but rather the experience of contributing to a 'good cause' (i.e. moral satisfaction), such that stated preference (SP) responses reflect 'warm glow' rather than true economic value. Elsewhere proponents of SP argue that insensitivity may be an artefact of study design issues (Carson and Mitchell, 1995), symbolic bias, questionable probability of provision¹, or due to mental models of joint products (Schulze et al 1998). Yet, some

¹ See, for example, Powe and Bateman (2004) who provide empirical evidence in support of the proposition that insensitivity may be result of differences in the perceived probability of provision.

apparent occurrences of scope insensitivity may be consistent with theoretical expectations (Bateman et al 1997). Economic theory tells us that WTP for a good will decline as the availability of substitute goods increases. In line with this, the value assigned to a good, comprising close substitutes, is expected to be less than the sum of the independently valued component parts (Hoehn and Randall, 1989). For example, Hoehn (1991) provides empirical evidence of how programmes to improve air quality in different regions of the United States result in lower WTP when jointly provided, compared with independent valuations. Insensitivity to scope is also expected in the face of preference satiation. For example, Rollins and Lyke (1998) illustrate how the creation of successive protected areas generated positive, but diminishing WTP, (e.g. marginal WTP for the first protected area was significantly greater than for the tenth area), so that sensitivity to scope was dependent upon where along the marginal demand curve the valuation was being conducted. The authors conclude that at the far right of the WTP curve scope sensitivity can be difficult to detect (and may require the use of large sample sizes).

Thus, insensitivity to scope may or may not be indicative of theoretically inconsistent choice behaviour. Moreover, in practise, it may not be straightforward to distinguish between the two situations. The detection of invalid WTP responses typically relies on an assessment of self-reported responses to follow-up questions, yet there may be manifold reasons for a respondent's reported WTP, including theoretically valid and invalid motives (Jorgensen and Syme, 2000). For example, Chilton & Hutchinson (2000) find that while warm glow may be a motivating factor in WTP, it does not preclude sensitivity to scope – that is, people may be 'impurely altruistic', such that they care about quality/quantity but also about moral satisfaction. Thus, the decision as to whether 'insensitive' responses are the result of inconsistent preferences rather than rational economic behaviour may not always be a clear one, and may require the design of additional tests – or more specific dis-embedding questions (see for example, Schulze et al 1998).

The usefulness of the test is further cast into doubt by findings in behavioural economics which suggest that insensitivity to scope is not just a hypothetical outcome but can be an "ordinary economic phenomenon" (p.370, Randall and Hoehn, 1996). If this is the case then the scope test cannot be meaningfully used as evidence of inconsistent preferences; rather it may simply indicate that Hicksian theory is insufficient (Bateman et al 1997).

In summary, there are various challenges associated with the use of scope tests as indicators of valid SP results. While there is strong basis for expecting a larger WTP for more of a good, such sensitivity to scope cannot be regarded as a "necessary nor sufficient" indicator of preference consistency (Bannerjee and Murphy, 2005: 613). These findings highlight the importance of developing better testing methods if the results of stated preference studies are to be reliably used in cost benefit analysis (CBA). In particular tests are required which can provide clear expectations with regard to theoretical consistency.

3 RESEARCH DESIGN AND HYPOTHESES

In accordance, our objective was to assess the consistency of preferences for tropical wildlife conservation using much more rigorous tests than those previously applied. As well as using scope tests we take lessons from experimental economics where we have clear *a priori* expectations about the outcomes we would expect in the presence of well-behaved preferences. Specifically, we examine the sensitivity of preferences to 'substitution effects' by varying the degree of similarity of the component parts of a more inclusive good. Further, we examine the robustness of preferences against anomalies by testing for ordering effects. We do so in the context of an online choice experiment to elicit the preferences of UK residents for conserving wildlife in the Eastern Arc Mountains.

The Eastern Arc Mountains are a chain of mountains covered in tropical forest and grasslands, which run the length of Tanzania. They are globally important for biodiversity, forming part of the Afrotropical hotspot, as well as being an important ecoregion in its own right (Burgess et al. 2004, Burgess et al. 2006). The forest areas provide critical habitat for a large number of endemic species, as well as many other rare and endangered wildlife. However these areas are under continued pressure from conversion and today less than 30% of the original forest remains.

The conservation of such wildlife may generate significant existence value to populations in 'donor' countries like the United Kingdom. However, the elicitation of such values is particularly challenging. The good itself is highly complex and unlikely to be familiar to respondents, raising concerns that preferences may be shaped by the valuation process rather than in relation to the characteristics of the good (Spash and Hanley, 1995)². The scant literature on the preferences of distant beneficiaries for conserving tropical biodiversity, provides some evidence of such effects: Horton et al (2003) in eliciting preferences of UK and Italian residents for conserving Brazilian Amazon (using responses to 'first seen' task) find that in external tests UK residents were sensitive to scope while Italian residents were not, (in both cases responses were highly variable). Further, Svendsater (2000), found that neither contingent valuation nor categorical rating were able to ensure respondents sensitivity to scope in external tests of four different types of global environmental good, including rainforests and endangered species.

To test the theoretical consistency of preferences in the current context we use a split-sample design comprising four treatments (see Table 1). In all treatments, respondents were required to answer two choice experiment tasks, one focusing on the conservation of wildlife in the Eastern Arc Mountains *only* (EA), and, the other on conservation of wildlife in the bundled Eastern Arc Mountains *and* the Cameroon Highlands ('EA+CH')³. The latter served the dual role of presenting the primary location of interest (the Eastern Arc) alongside a substitute (or complimentary) investment opportunity, and provided the more inclusive good for the scope test (i.e. wildlife conservation in the two locations bundled together).

Table 1: Research design

Treatment	Ordering Effect		Substitution
	1 st Task	2 nd Task	
A	EA	EA & CH	Similar species
B	EA & CH	EA	
C	EA	EA & CH	Dissimilar species
D	EA & CH	EA	

² It is oft argued that familiarity and experience are essential in obtaining reliable and meaningful value estimates for environmental goods, to the extent that individuals cannot be expected to hold stable and well-formed preferences for complex and unfamiliar goods (Spash, 2007; Kahneman & Knetsch, 1992; Gregory et al 1993). Indeed, Boyle et al (1993) find that ordering effects are more likely amongst respondents unfamiliar with the good.

³ The Cameroon Highlands were selected due to its broad similarity with the Eastern Arc (they comprise a range of mountains covered in patches of rainforest and grasslands also located in Africa, and form part of the Guinean Forests biodiversity hotspot) enabling us to control for factors exogenous to our CE design, such as location and ecosystem type.

To test for procedural invariance the presentation order of the two choice tasks was varied across treatments. They were framed as an *exclusive* list (Bateman et al 2004); meaning goods were presented as alternatives rather than additions to the previously seen good⁴. This is an important point, since it means that individual reference levels of income, utility, prices and so on, remain constant across choice tasks, such that any difference in WTP may be regarded as the outcome of an ordering effect (Bateman et al 2004).

A novel aspect of our design was to investigate the potential impact of substitution effects on preferences (in particular, scope sensitivity) which we accomplished by varying the level of similarity of the wildlife at risk in the Cameroon Highlands, so that these were broadly similar and dissimilar to those contained in the Eastern Arc.

The wildlife conservation in the Eastern Arc was described by three attributes: (i) the number of '*unique*' species saved/extinct in the *Eastern Arc*; (ii) the number of '*non-unique*' threatened species saved/lost in *Eastern Arc*; and, (iii) the annual donation required from the household⁵. Attribute levels for the Eastern Arc were based on expert knowledge of current biodiversity levels in the area and expectations under three different future scenarios – which were part of the wider project goals. The species most likely to be affected comprised an endemic lizard, two endemic bird species and ten endemic species of frogs/toads, and the non-endemic African Lion. The levels selected for the cost variable were based on literature review and pre-testing.

In the bundled 'EA+CH' task, a further two attributes were included to describe: (iv) the number of '*unique*' species saved/extinct in *Cameroon Highlands*; and, (v) the number of '*non-unique*' species saved/extinct in *Cameroon Highlands*. Attribute levels for the Cameroon Highlands were manipulated to meet the purposes of the experimental design⁶. In the similar treatment, the species at risk in Cameroon Highlands comprised an endemic lizard, an endemic bird and six endemic frog/toad species and the African Lion, but in the dissimilar treatment the unique CH lizard was replaced by a unique gorilla species, and the CH non-unique lion was replaced by a non-unique frog. A description of the attributes and levels are set out in Table 2.

In both choice tasks, the experimental design was based on the full factorial⁷, with implausible options removed⁸. In the case of the EA task, this generated a total of 30 choice

⁴ In treatments where the Eastern Arc only was presented first, the pre-text to the second choice task explained to respondents that *instead* of focusing only on the Eastern Arc, WWF could also allocate funding to the Cameroon Highlands; information was then presented on the Cameroon Highlands. Likewise, in treatments where the bundled Eastern Arc and Cameroon Highlands (EA & CH) were presented first, the pre-text to the second choice experiment task explained to respondents that *instead* of targeting both sites, WWF could instead focus on the Eastern Arc *only* (EA). In this way, the nested goods (eastern arc and bundled eastern arc and Cameroon highlands) were presented as an *exclusive* list (Bateman et al 2004).

⁵ An annual donation was adopted since these mechanisms are commonly used by international conservation organisations as a way to capture public willingness to pay for wildlife conservation efforts (eg 'adopt a lion' schemes), correspondingly, it was thought to offer a realistic and credible payment mechanism (supported by the results of cognitive testing), while avoiding potential protests associated with compulsory mechanisms such as taxes.

⁶ Efforts were made to ensure the levels used were broadly in keeping with the bio-geography of the site, but we cannot claim they represent current or future biodiversity levels in the Cameroon Highlands with any accuracy.

⁷ This allows estimation of all possible interaction effects and seemed the most appropriate design option given no clear *a priori* expectations of the nature of interactions (the latter being required for optimal designs).

⁸ In each of the choice tasks, six implausible options were identified (options which combined status quo levels with a positive monetary payment).

cards, which were randomly allocated into five blocks of six choice questions⁹. In the 'EA+CH' choice task, a total of 210 choice cards were generated, which were randomly allocated into thirty blocks of seven choice questions. The survey was programmed so that respondents were randomly allocated to a treatment, and blocks were randomly selected with equal probability of each EA block being combined with each 'EA+CH' block. In addition, choice questions were presented in a random order to avoid any possible internal sequencing effects.

Table 2: Attributes and Levels

Attributes	Levels	
	Eastern Arc (EA)	Eastern Arc & Cameroon Highlands (EA & CH)
EA Unique Sp	13 sp saved	13 sp saved
	12 sp saved	12 sp saved
	0 sp saved (SQ)	0 sp saved (SQ)
EA Non-uniq Sp	Lion saved	Lion saved
	Lion lost (SQ)	Lion lost (SQ)
CH Unique Sp	-	8 sp saved (incl lizard ^{A/b} /gorilla ^{C/d})
	-	7 sp saved
	-	0 sp saved (SQ)
CH Non-uniq Sp	-	Lion ^{a/b} /Frog ^{c/d} saved
	-	Lion ^{a/b} /Frog ^{c/d} lost (SQ)
Annual Donation	£0 (SQ), £5, £20, £40, £60, £100, £200	

Notes: EA = Eastern Arc Mountains; CH = Cameroon Highlands; Sp = species; SQ = Status quo. Superscripts ^{a, b, c, d} indicate the experimental treatment: in the similar treatments (A & B) the CH 8 unique species include lizard, frog, birds and non-unique is lion; and in the dissimilar treatments (C & D), the CH 8 unique include gorilla, frog, birds and non-unique is frog.

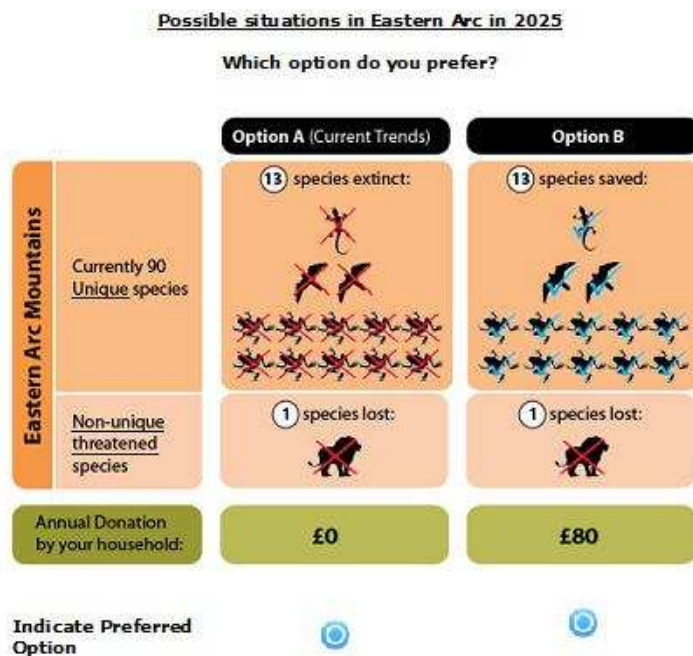
To this end, each respondent was required to answer six choice questions on the Eastern Arc only and seven choice questions on the bundled Eastern Arc and Cameroon Highlands. The contingent scenario explained to respondents that wildlife in the Eastern Arc (or Eastern Arc and Cameroon Highlands) was under threat from loss of habitat, and that WWF was working with local governments to reduce species losses by setting up protected areas and so on. It was further explained that WWF would require funding from members of the public if the programmes were to go ahead - the purpose of the research being to find out which schemes people would be prepared to fund and which schemes they would not fund. In each

⁹ Using the random number generation command in excel each card was randomly allocated to a block, the resulting design matrix (including the blocking number) was subjected to visual inspection, tested for possible correlations, and adjusted until a minimal level of correlation was achieved.

task, respondents were presented with a choice between two options: the status quo and a conservation alternative.

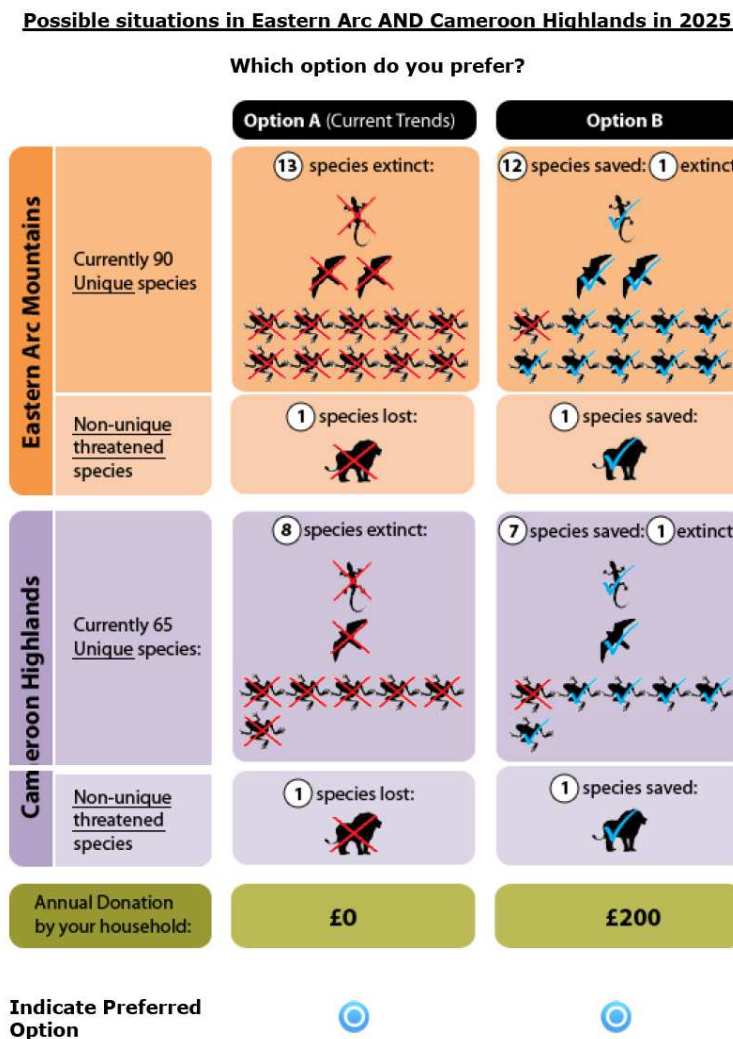
A further important aspect of our design involved the use of symbols (as well as numbers) to convey to respondents the number and type of animals affected under the various options in the choice cards. This follows the findings of Bateman et al (2009) that numeric information may lack the ‘evaluability’ of visual representations of the same data, (in their case using a virtual reality experiment to value land use change)¹⁰, and was facilitated through the use of an online design. The emergence of internet or online surveys provides a promising opportunity to enhance the information presented to respondents through visual stimuli, colour, and higher quality images. (In our case, it also offered a cheaper option for reaching respondents across the UK with relative ease - the wider purpose of our study being to estimate aggregate UK benefits.). An example of an EA choice card and a *similar* ‘EA+CH’ choice card are presented in Figures 1 and 2, respectively.

Figure 1: Example choice card used in the EA choice experiment



¹⁰ Christie et al (2006) also find that visual information transfer approaches (in their case a power point presentation) are more suitable for presenting biodiversity which was found to be unfamiliar and considered complex by respondents.

Figure 2: Example choice card used in the 'EA & CH' choice experiment



3.1 Hypotheses:

To examine external sensitivity to scope¹¹ we undertake a test of *categorical* nesting (Carson and Mitchell, 1995), by comparing the welfare estimates for the 'Eastern Arc' and the bundled 'Eastern Arc and Cameroon Highlands' (similar), using responses to first valuation task only.

Theory tells us that WTP should be larger for more of a good, thus, if estimates are scope sensitive then we would expect that WTP for the part (EA) is less than WTP for the more inclusive good (EA+CH):

$$H^0: WTP^1_s (EA) < WTP^1_s (EA+CH) \quad [1.1]$$

where the superscript 1 means the good is presented first and the subscript s means the component parts are similar. Moreover, we would also expect that WTP for the small good

¹¹ Internal scope tests are considered to be a relatively weak indication of scope sensitivity as respondents may simply be trying to be internally consistent.

(EA) would be the same regardless of whether valued independently or embedded in the more inclusive good (EA+CH):

$$H^0: WTP^1_s (EA) = WTP^1_{s/e} (EA) \quad [1.2]$$

where subscript *e* indicates the good is embedded within the more inclusive good.

The above tests follow standard approaches reported in the literature for assessing scope sensitivity. As discussed, a shortcoming of many of these studies is that they fail to take into account the potential impact of substitution effects on scope sensitivity - critically, in the face of near perfect substitution coherent and well-behaved preferences may indeed appear insensitive to scope. To address this, we explicitly incorporate a test to examine the impact of substitution effects on scope sensitivity by varying the degree of nesting. We do so, by manipulating the level of similarity of the components parts of the more inclusive good so that the wildlife in EA and CH are now *dissimilar* and re-examine sensitivity of WTP for the part (EA) is less than for more inclusive good (EA+CH):

$$H^0: WTP^1_s (EA) < WTP^1_d (EA+CH) \quad [2.1]$$

where superscript *1* means the good is presented first and subscript *d* means the component parts are dissimilar. We also re-examine whether WTP for the small good (EA) is the same regardless of whether it is embedded in the *dissimilar* more inclusive good (EA+CH) or valued independently:

$$H^0: WTP^1_s (EA) = WTP^1_{d/e} (EA) \quad [2.2]$$

If we find evidence that substitution effects occur, then we can argue that preferences are unlikely to be arbitrarily constructed.

We further test the theoretical consistency of preferences by examining the robustness of welfare estimates to order effects. We do so using an *exclusive* list since this provides clear *a priori* expectations with regard to theoretical consistency (Bateman et al, 2004)¹². Economic theory states that the value of a good in an exclusive list should be invariant to its position within the list. A number of empirical studies have found evidence to the contrary. In the current context, to formally test for ordering effects, we would expect that for procedural invariance to hold:

$$H^0: WTP^1_s (EA) = WTP^2_s (EA) \quad [3.1]$$

$$H^0: WTP^1_s (EA+CH) = WTP^2_s (EA+CH) \quad [3.2]$$

$$H^0: WTP^1_d (EA) = WTP^2_d (EA) \quad [3.3]$$

$$H^0: WTP^1_d (EA+CH) = WTP^2_d (EA+CH) \quad [3.4]$$

Where superscripts *1* and *2* indicate the presentation order (1= 'first seen' task; and, 2 = 'second seen' task); and subscripts *s* and *d* indicates goods in treatment are *similar*, and *dissimilar*, respectively. The occurrence of an 'ordering effect' would violate the requirement for procedural invariance, thereby indicating the presence of inconsistent preferences.

4 ECONOMETRIC MODEL

The latent unobservable utility U_{ai} for choice alternative *a* held by individual *i* can be described in terms of its observable component V_{ai} and its unobservable component ε_{ai} , the error term:

¹² As opposed to Kahneman and Knetsch (1992) who use inclusive lists.

$$U_{ai} = V_{ai} + \varepsilon_{ai} \quad [1]$$

This expression can be disaggregated further to show that utility is assumed to be a function of the characteristics of a good (Z_{ai}) and in addition may also vary according to characteristics of the individual (S_{ai}):

$$U_{ai} = V(Z_{ai}, S_{ai}) + \varepsilon(Z_{ai}, S_{ai}) \quad [2]$$

We use a binary choice experiment, where respondents are asked to choose the alternative $a = \{0, 1\}$ which maximises their utility given a choice set with two alternatives which differ by a particular set of k attributes, $Z_i = \{z_{1i}^a, z_{2i}^a, \dots, z_{ki}^a\}$. The probability that individual i prefers alternative 1 over the base alternative is given by:

$$\Pr(\theta_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)} \quad \text{where, } \theta_i = \sum_{k=1}^K \beta_k (z_{ki}^1 - z_{ki}^0). \quad [3]$$

Given that individuals are required to answer more than one choice set, responses per individual may be correlated across choice sets (although not across individuals). In order to account for this we use the *random effects* logit model which allows error terms to be freely correlated for each individual but uncorrelated across individuals, so that equation 1 now becomes:

$$U_{ait} = V_{ait} + (\varepsilon_{ait} + v_i) \quad [4]$$

where t indicates the group of choices made by individual i and v_i is the unobserved individual-specific heterogeneity. Following maximum likelihood estimation, welfare measures for alternative a can be estimated using:

$$WTP_i^a = \frac{-1(\beta_1^a z_{1i}^a + \beta_2^a z_{2i}^a + \dots + \beta_k^a z_{ki}^a)}{\beta_{price}^a} \quad [5]$$

and, the marginal value (or implicit price) of a change in any single attribute is given by:

$$WTP = \frac{-\beta_{attribute}^a}{\beta_{price}^a} \quad [6]$$

where β_{price}^a is the coefficient on the price attribute.

5 SURVEY IMPLEMENTATION AND SAMPLE QUALITY

The survey was administered by Survey Sampling International. The sample was selected using a stratified approach aimed at eliciting a nationally representative sample from a pre-recruited panel of online UK residents. An email invitation containing a web-link to the survey instrument was sent to 42,264 residents, across the UK, in January/February and April/May, 2009; a reminder email was sent to those who had not responded after 36 hours. To control access to the survey, respondents were given a unique ID and password which enabled them to complete the survey only once, but allowed them to carry on where they left off if they were unable to complete in one go. Survey responses were transmitted via http, stored in databases, then exported using Nebu software to SPSS and excel.

In total, 6065 individuals started the survey¹³, of these around 45% were screened out due to quota satiation, a further 20% dropped out and a further 35% completed the survey. After adjustments to ensure representativeness, equivalence between treatments, and censoring for invalid and protest responses¹⁴, a final sample of 999 complete questionnaires was generated, approximately evenly split between the four treatments.

An overview of the key socio-demographic characteristics of each sub-sample is provided in Table 3. No significant differences were detected for gender, age, household income and education between treatments. The sub-samples were found to be broadly ‘representative’ of the UK population¹⁵.

Table 3: Key socio-demographics across treatments (censored sample)

	A	B	C	D	UK
Males (%)	45.3%	46.8%	45.4%	46.6%	48%
Av. Age (mid-pt)	46.9	47.4	46.7	46.2	-
Educ: degree level or above	21.1%	22.0%	20.3%	20.3%	20%
Av. HsH income/yr (mid-pt)	£31,974	£31,960	£32,162	£31,843	-
Sample size (N)	247	250	251	251	

6 ECONOMETRIC RESULTS

The choice responses were analysed using a logit model with random effects. The model for the Eastern Arc [7] is specified as:

$$V_{it}EA = \beta_1 Const_{it} + \beta_2 EA_13uniq_{it} + \beta_3 EA_12uniq_{it} + \beta_4 EA_lion_{it} - \beta_5 Cost_{it} + \varepsilon_{it} \quad [7]$$

and the model for the bundled Eastern Arc and Cameroon Highlands [8] are set out below:

$$V_{it}(EA + CH) = \beta_1 Const_{it} + \beta_2 EA_13uniq_{it} + \beta_3 EA_12uniq_{it} + \beta_4 EA_lion_{it} + \beta_5 CH_8uniq_{it} + \beta_6 CH_7uniq_{it} + \beta_7 CH_Lion / Frog_{it} - \beta_8 Cost_{it} + \varepsilon_{it} \quad [8]$$

where V_{it} is the vector of the option variable for individual i in group t , the β s are the estimated coefficients of the variables, and ε_{it} is the unobserved component of utility different in eq. 7 and 8.

The estimation results are presented in Table 4 – the top half of the table presents results for the ‘first seen’ choice task and the bottom half the ‘second seen’ task. In the following

¹³ This is comparable with other internet-based SP surveys which typically have much lower response rates compared with other survey modes, for example, Marta-Pedroso et al (2007) report response rates of 5.1% for internet and 84% for in-person.

¹⁴ Invalid responses were identified as those given by respondents that selected the ‘do something’ option in all choices for non-economic reasons, for example, ‘I get a sense of satisfaction from contributing to good causes’, ‘I wouldn’t really pay’ or ‘We have a moral duty’. Protest responses were identified as those given by respondents that selected the ‘do nothing’ (i.e. status quo) option in all choices for non-economic reasons, for example, ‘The government should pay for this’, ‘I don’t believe the money would really be spent on the programme’ or ‘The choices were too difficult’. In this way, a total of 33 respondents were excluded for invalid positive responses and a further 38 respondents were excluded for protest responses.

¹⁵ Internet surveys may struggle with providing representative samples. This reflects the fact that not all households have internet access and access is more likely among certain households than others (Berrens et al 2004). For example, in 2008, 16.46 million UK households had Internet access, representing 65% of UK households (ONS, 2008).

discussion we refer to the results for the 'first seen' tasks only in order to avoid any unintended ordering effects – the latter are discussed in subsequent sections.

The model results for the small good (Eastern Arc), indicate that all attributes are significant and have the expected signs. This provides some weak indication of sensitivity to scope: respondents are willing to pay positive amounts to secure more of a good. On closer inspection, we note that the coefficient on saving 13 unique species is slightly smaller than that for saving 12 unique species –however, results of Wald test indicate that the difference is not statistically significant (chi2 stat of 2.04; p value 0.153). Thus, WTP for conserving 12 unique species is statistically equal to WTP for conserving 13 species. Given we are at the maximum protection levels, it certainly seems plausible that the apparent insensitivity between 12 and 13 species may be due to diminishing marginal utility (Rollins and Lyke, 1998), yet it remains the case, that many other explanations may also fit. For example, after Schulze et al (1998), respondents may believe that in paying to conserve 12 species, the whole area will be conserved, thereby protecting all thirteen endangered unique species. Alternatively, this may be a manifestation of 'warm glow'. This highlights one of the difficulties in interpreting the results of scope tests – how to determine whether apparent insensitivity is the result of preferences which are consistent with economic theory (and can be explained by, for example, satiation or substitution effects) or whether scope insensitivity can be attributed to reasons that are more detrimental to stated preference methods (such as the presence of arbitrarily constructed preferences, warm glow effects etc).

The results for the large 'similar' good (where the Eastern Arc is bundled with Cameroon Highlands and this contains similar species), indicate that all attributes are correctly signed and significant, with the exception of the coefficients on the 'Eastern Arc lion' and '7 unique species in Cameroon'. The insignificance of the 'Eastern Arc lion' is an interesting outcome given the coefficient on the 'Cameroon Highlands lion' is significant and positive. This suggests that some kind of substitution effect takes place when the lion is provided in both locations i.e. preferences are satiated so long as the lion is conserved in at least one location, in this case the Cameroon highlands. The reason for the insignificance of the '7 unique species in Cameroon highlands' is more complex. One possible explanation is that it may simply be that conserving '7 unique species' is not enough to significantly contribute to utility i.e. there is a minimum bound on what people are prepared to pay for.

The results for the large 'dissimilar' good are analogous: again all attributes have the expected sign and almost all are significant, however, in this treatment whereas we observe that the 'Eastern Arc lion' is almost significant, the 'Cameroon highland non-unique frog' is highly insignificant. This seems to corroborate our earlier finding of a possible substitution effect between the two lions; and also highlights the 'charisma' effect that such a mega fauna may have. When the lion in the Cameroon Highlands is replaced by a frog; the non-unique attribute becomes highly insignificant – this is in accord with findings from the literature that has shown that people are typically WTP more for charismatic or iconic species (see for example, Loomis and White, 1996; Kontoleon and Swanson 2003). Further, we see that when the lion is found in the Eastern Arc only (and not in the Cameroon Highlands), the coefficient on this attribute increases greatly and is very close to being significant - note, in this treatment the Cameroon Highlands contains a 'unique' gorilla species which may have an additional effect on the significance of the lion.

In summary, for the most part we observe internal sensitivity to scope while apparent insensitivity seems to be attributable to substitution effect. This, conjecture is further examined in the subsequent sections.

Table 4: Regression Results for small good (EA) and large good (EA+CH): (i) 1st presented and (ii) 2nd presented across treatments (censored)

1st presented	small good^{A/C}				large good (similar)^B				large good (dissimilar)^D			
Variables	coeff.	se	t stat	p	coeff.	Se	t stat	p	coeff.	se	t stat	p
Constant	0.38	0.20	1.94	0.05	0.48	0.22	2.16	0.03	0.40	0.23	1.74	0.08
EA Uniq 13 sp	0.65	0.14	4.77	0.00	0.39	0.15	2.60	0.01	0.90	0.17	5.32	0.00
EA Uniq 12 sp	0.80	0.12	6.53	0.00	0.36	0.14	2.55	0.01	0.52	0.14	3.76	0.00
EA Non-uniq lion	0.38	0.11	3.46	0.00	0.09	0.12	0.75	0.46	0.16	0.11	1.45	0.15
CH Uniq 8 sp					0.29	0.15	1.94	0.05	0.72	0.17	4.35	0.00
CH Uniq 7 sp					0.17	0.14	1.17	0.24	0.53	0.15	3.51	0.00
CH Non-uniq: lion ^{a/b} /frog ^{c/d}					0.19	0.11	1.72	0.09	0.04	0.12	0.30	0.76
Cost	-0.02	0.00	-21.12	0.00	-0.02	0.00	-17.42	0.00	-0.02	0.00	-16.76	0.00
Rho	0.78	0.01	58.71	0.00	0.77	0.02	45.51	0.00	0.80	0.02	49.04	0.00
Correct Predictions	65.6%				65.1%				64.3%			
Log Likelihood	-1376.5				-802.7				-781.9			
Obs.	498				250				251			

2nd presented	small (similar)^B				small (dissimilar)^D				large (similar)^A				large (dissimilar)^C			
Variables	coeff.	se	t stat	p	coeff.	Se	t stat	p	coeff.	se	t stat	p	coeff.	se	t stat	p
Constant	0.70	0.27	2.56	0.01	0.65	0.28	2.34	0.02	0.68	0.22	3.16	0.00	0.98	0.23	4.25	0.00
EA Uniq 13 sp	0.66	0.19	3.53	0.00	1.04	0.19	5.44	0.00	0.67	0.20	3.41	0.00	0.80	0.18	4.42	0.00
EA Uniq 12 sp	0.49	0.17	2.80	0.01	0.78	0.19	4.06	0.00	0.47	0.18	2.65	0.01	0.72	0.17	4.16	0.00
EA Non-uniq lion	0.18	0.15	1.23	0.22	0.56	0.15	3.83	0.00	0.03	0.13	0.20	0.84	0.54	0.15	3.64	0.00
CH Uniq 8 sp									0.57	0.19	3.07	0.00	1.01	0.21	4.95	0.00
CH Uniq 7 sp									0.60	0.19	3.13	0.00	0.38	0.19	1.97	0.05
CH Non-uniq: lion ^{a/b} /frog ^{c/d}									0.33	0.13	2.49	0.01	0.36	0.15	2.40	0.02
Cost	-0.02	0.00	-15.36	0.00	-0.02	0.00	-17.35	0.00	-0.03	0.00	-15.83	0.00	-0.03	0.00	-15.39	0.00
Rho	0.81	0.02	43.71	0.00	0.82	0.02	49.35	0.00	0.86	0.02	59.15	0.00	0.89	0.01	64.89	0.00
Correct Predictions	65.6%				64.9%				63.7%				63.8%			
Log Likelihood	-676.9				-674.6				-709.2				-698.2			
Obs.	250				251				247				251			

Note: superscripts ^{a, b, c, d} indicate treatments. In similar large: CH 8 unique species include lizard, frog, birds and non-unique is lion; in dissimilar treatment, 8 uniq include gorilla, frog, birds and non-unique is frog.

7 PRELIMINARY OBSERVATIONS ON WELFARE ESTIMATES

Predicted mean welfare measures (and 95% confidence intervals) for the small and large goods across the various treatments are reported in Table 5 and further presented in Figure 3. Initial observations indicate welfare estimates are characterised by wide (and in most cases, overlapping) confidence intervals, suggesting high variability in responses. This is consistent with other studies of WTP for tropical biodiversity amongst distant beneficiaries (see Horton et al 2003). And, may also be an artefact of the survey implementation mode (recent research suggests that online surveys may actually result in poorer data quality due to the more interactive stimuli placing greater cognitive burden on respondents - see Savage and Waldman, 2008). Only in extreme cases do we observe significant differences in welfare estimates: small similar good presented 2nd (Trt B) vs. large dissimilar good presented 2nd or 1st (Trt C or D) at the 95% confidence interval; and small similar presented 1st (Trt A/C) vs large dissimilar presented 2nd (Trt C) at the 90% confidence interval. Nevertheless, comparing point estimates of mean WTP, there does appear to be some coherency in responses: WTP for small good is consistently lower than WTP for large good, suggesting internal sensitivity to scope. And, in line with our expectations, we find that mean WTP for the large 'similar' good tends to be lower than WTP for large 'dissimilar good'. We expect that with a larger sample size (or face-to-face surveys rather than an online survey) the expected differences between welfare measures may become statistically significant. We examine these trends in greater detail in the subsequent sections.

Table 5: Mean WTP for small and large good across treatments (for conserving all wildlife)

SIMILAR		DISSIMILAR	
Small	Large	Small	Large
£52.01 ^{*A/C} (£31.28-£72.74)	→ £64.63 ^A (£40.07-£89.19)	£52.01 ^{*A/C} (£31.28-£72.74)	→ £87.64 ^C (£66.72-£108.56)
£38.99 ^B (£15.27-£62.71)	← £54.12 ^B (£23.77-£84.47)	£71.45 ^D (£45.37-£97.53)	← £91.58 ^D (£61.52-£121.64)

Note: superscripts indicate treatment sub-sample; arrows indicate order of valuation tasks; 95% confidence intervals are in parentheses; * based on pooled sub-samples for treatments A and C.

Figure 3: Pairwise comparisons of welfare measures based on 95% confidence intervals

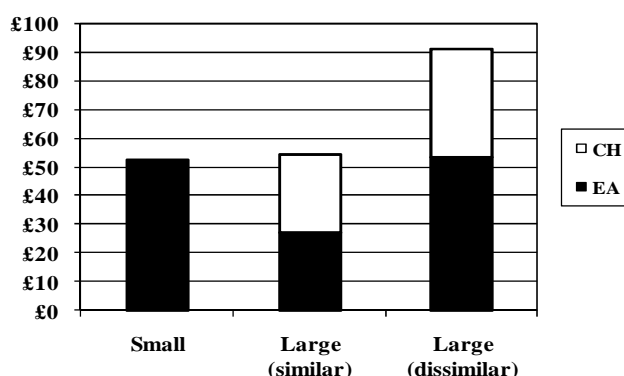
	1 st	2 nd	2 nd	2 nd	1 st	2 nd	1 st	
	EA ^{A/C}	EA ^B	EA ^D	EA + CH ^A	EA + CH ^B	EA + CH ^C	EA + CH ^D	
1 st	EA ^{A/C}							Sim
2 nd	EA ^B	=						Sim
2 nd	EA ^D	=	=					Dissim
2 nd	EA + CH ^A	=	=	=				Sim
1 st	EA + CH ^B	=	=	=	=			Sim
2 nd	EA + CH ^C	(≠)	≠	=	=	=		Dissim
1 st	EA + CH ^D	=	(≠)	=	=	=	=	Dissim
		Sim	Sim	Dissim	Sim	Sim	Dissim	Dissim

Notes: parenthesis indicate significance differences based on 90% confidence intervals

8 TESTING FOR EXTERNAL SENSITIVITY TO SCOPE

Our results indicate that on average respondents are WTP per annum £52.01 and £54.12 for conservation of the Eastern Arc only and for the bundled Eastern Arc and the Cameroon Highlands where the components are *similar*, respectively (see first two columns in Figure 4). Taken on its own, this would support a finding of insensitivity to scope: respondents are willing to pay virtually the same amount for a small good as for a more inclusive similar good.

Figure 4: Mean WTP for small good, large (similar) and large (dissimilar) goods



To further explore this, we compare the welfare estimates for the small good (EA) when valued on its own and as *part* of a more inclusive *similar* good (shaded areas in Figure 4). Interestingly, mean WTP for the Eastern Arc is around 50% lower when valued as 'part' of a more inclusive *similar* good compared with being valued on its own, £52.01 and £27.12, respectively. This further suggests the presence of a strong embedding effect for the small similar good.

Table 6: welfare measures for Eastern Arc when valued on its own ('whole') and as a 'part'

	Whole	Part (similar)	Part (dissimilar)
EA	£52.01	£27.12	£53.22
	(£31.28-£72.74)	(£5.86-£48.39)	(£33.66-£72.79)

Notes: parentheses contain 95% confidence intervals

However, there are a number of possible explanations for these findings (indeed this is precise problem with scope tests). They could be the result of Kahneman and Knetsch's (1992) moral satisfaction (i.e. people are not valuing the good itself but rather the act of giving to a good cause, consequently WTP is invariant to a change in the scope of a good). Alternatively, respondents preferences may simply be satisfied by conserving the small good (EA) so that adding more species conservation in another location does not significantly increase utility i.e. after Rollins and Lyke (1998) they are at the far right of the marginal WTP curve. Or, it may be that we are simply observing a substitution effect (Carson & Mitchell 1995 and Hoehn & Randall, 1989) - if species contained in EA are perceived to be near perfect substitutes with those contained in CH, then wildlife conservation in EA may be valued the same as wildlife conservation in EA & CH.

9 INCORPORATING SUBSTITUTION EFFECTS IN TESTING FOR EXTERNAL SENSITIVITY TO SCOPE

To investigate the possibility that substitution effects are causing this apparent scope insensitivity, our design manipulates the type of species present in CH so that in the large ('EA+CH') *dissimilar* good the component parts contains dissimilar wildlife.

The results indicate that contrary to the findings for the large *similar* good, mean WTP per annum for the large *dissimilar* good is £91.58, almost twice as much as WTP for the small good (EA), £52.01 (see third column in figure 4). That is, WTP for the small good is virtually identical to WTP for the more inclusive good when the latter comprises close substitutes; yet when we manipulate the species contained in the more inclusive good so that there is more variability (i.e. less substitution between component parts), WTP is almost twice as much as for the small good. This is consistent with a finding of substitution effects. Crucially, it suggests that respondents are behaving in a consistent manner: they are indifferent between conserving one hotspot or two hotspots when these contain the same kind of wildlife, but when the two hotspots contain very different animals, WTP almost doubles for the conserving the additional hotspot (respondents are WTP more for more variability).

Furthermore, in relation to the earlier finding of an apparent 'embedding effect' for the small *similar* good (EA) - we observe no such effect when the small good is valued as part of a *dissimilar* more inclusive good. In fact, mean WTP per annum for the EA when valued as part of the dissimilar large good is £53.22, almost identical to when valued on its own, £52.01, respectively (shaded part in figure 4). These results suggest that the observed scope insensitivity is actually the result of respondents adjusting their WTP for the Eastern Arc in response to the presence or absence of close substitutes. This supports our earlier supposition that while respondents may appear to be insensitive to scope when the goods are similar, this apparent 'embedding effect' may simply mask a 'substitution effect'.

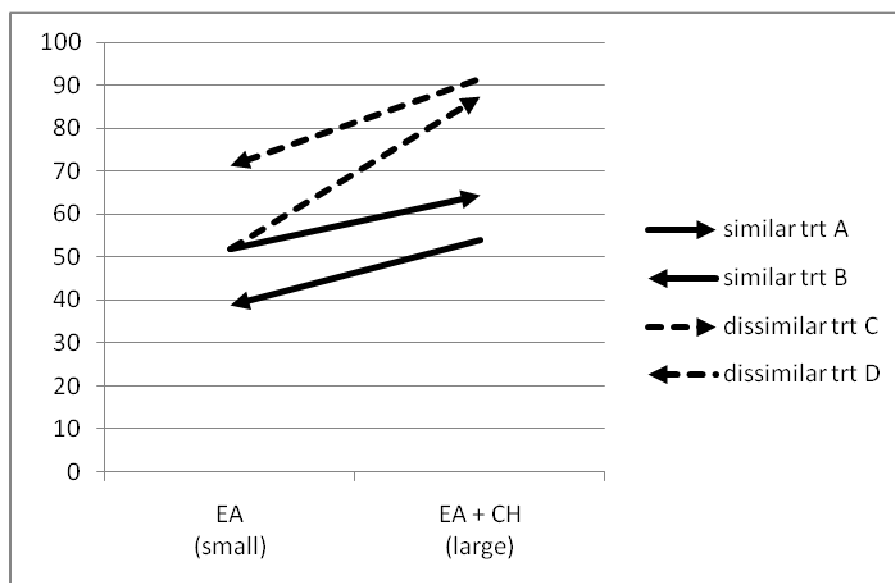
10 TESTING FOR 'ORDER EFFECTS'

To further test the consistency of preferences for tropical wildlife we undertake tests to examine robustness towards ordering effects. The results are presented in Figure 5 (detailed descriptions of welfare estimates are provided in Table 5).

A number of preliminary observations seem striking. With respect to the *similar* treatments, the slope of the WTP curve is virtually identical regardless of whether we move from bottom-up (small to large) or top-down (large to small). In other words, respondents seem to adjust their WTP for the *same* change by very similar amounts regardless of the presentation order, this suggests some consistency in preferences in the *similar* treatments. In contrast, in the *dissimilar* treatments, the slope of the curves is very different: respondents appear to be willing to pay a lot more for the small good when seen after the large good compared with beforehand. That is, they adjust their WTP for the same change by different amounts depending on whether they see the large or small good first.

A closer inspection of results reveals mean WTP per annum for the small good (Eastern Arc) is £52.01 when presented as the first choice task, £38.99 when presented after a *similar* large good and £71.45 when presented after a *dissimilar* large good. In all cases, confidence intervals overlap, indicating that we cannot reject the null hypotheses of procedural invariance for the small good.

Figure 5: Mean WTP for small and large scope goods across treatments



Notes: arrows indicate the direction of valuation task

With respect to the large good (bundled EA and CH), when the component parts are similar, mean WTP per annum is £54.12 when presented as the first choice task and £64.63 when presented as the second task (i.e. after the small similar good). When components are dissimilar, mean WTP per annum is £91.58 when presented as the first choice task and £87.64 when presented as the second task after the small good. Consistent with findings for the small good, the 95% confidence interval overlap indicating that welfare measures for the more inclusive good do not appear to be affected by ordering effects. Based on these results, we fail to reject hypotheses [3.1] to [3.4] and conclude there does appear to be some evidence that preferences are invariant to the presentation order.

To further investigate these findings, we ran likelihood ratio tests to examine overall model equivalence for EA and 'EA+CH' models based on presentation order. Results confirm that taste parameters for the small good are equivalent whether seen before or after a larger *similar* good (LR stat = 5.624; chi2 stat=12.6 at 95% level with 6 df), and likewise, whether seen before or after a larger *dissimilar* good (LR stat = 9.1828; chi2 stat=12.6 at 95% level with 6 df) – providing further evidence of procedural invariance for the small good. However, a closer examination of the regression results indicates some differences at the attribute level. In line with expectations the coefficient on 'EA lion' is found to be positive and significant when the small good is seen first, and likewise when the small good is seen after the larger *dissimilar* good (which only contains the lion in the EA), but it is not significant when the small good is seen after a larger *similar* good (which contains the lion in CH as well as EA). We speculate that this attribute level order effect may arise due to respondents seeking to maintain internal consistency with responses to the 'first seen' choice task (i.e. where the EA lion was not significant) – i.e. an 'anchoring effect'. An intuitive explanation is that by asking respondents in the first choice task to make explicit choices about conserving the lion in Eastern Arc *and* the Cameroon Highlands, they give greater consideration to the fact that the lion is not unique to the Eastern Arc (and thus may be substituted with locations elsewhere), so that in the second task the Eastern Arc Lion is still not considered to be important. This seems to fit with findings elsewhere that simple reminders about existence of substitute goods may not significantly affect welfare estimates (see for example, Loomis et al 1994, Ojea and Loureiro, 2009) but, approaches which have sought to stimulate more explicit consideration of substitutes appear to have been more successful (see, for example, Hailu et al, 2000; Neill, 1995).

With respect to the larger good, the results of a likelihood ratio test for overall model equivalence, indicate the presence of weak order effects in the similar treatments (LR stat 15.88; chi2 stat=14.7, 90% sig level with 9 df) and strong order effects in the dissimilar treatments (1% significance level: LR stat 26.8; chi2 stat=21.7 99% sig level with 9 df). That is, the 'EA+CH' models appear to be significantly affected by whether they are seen before or after the EA choice tasks, suggesting some kind of anchoring effect on the EA model. A closer inspection of the regression results for the large, *similar* good, indicate that the coefficient on 'CH unique 7 species saved' is not significant when this good is seen first, but it is highly significant (1% sig level) when presented after the small good – it is unclear why preferences differ in this manner, we simply note that an attribute level ordering effect seems to occur. We also observe attribute level differences for the large, *dissimilar* good: when presented first the coefficients on the non-unique CH frog and non-unique EA lion are not significant; however when it is presented second all attributes are significant. We cannot be sure as to the reason for this effect but again it seems to be the result of respondents seeking to be internally consistent in the second task (where a larger dissimilar good is presented), by maintaining the significance of attributes that were considered to be important in the first task (i.e. small good). We might speculate that such “construction” arises due to increased task complexity in the dissimilar treatment: in these choice tasks respondents are faced with more variety in choice options (since CH animals are now dissimilar to EA animals and have not been previously seen), thus they may be more likely to resort to heuristics in decision process – this is conjecture.

11 CONCLUSIONS

This is one of the first studies to undertake rigorous testing of the theoretical consistency of preferences for wildlife conservation against firm expectations.

A shortcoming of many studies testing for scope sensitivity is that they fail to take into account the potential impact of substitution effects. Indeed, using standard approaches our results suggest that preferences for tropical wildlife conservation are insensitive to scope and susceptible to embedding effects - indicating preferences are not coherent. However, when we lessen the likelihood of substitution by manipulating the type of wildlife present in our conservation sites so that it is dissimilar, we observe very different findings: willingness to pay for the more inclusive good is almost twice as much as for the small good, and the apparent embedding effect disappears. Crucially, these results suggest that respondents are behaving in a consistent manner: they are indifferent between conserving one hotspot or two hotspots when these contain the same kind of wildlife, but, are willing to pay almost double when the two hotspots contain very different animals (respondents are WTP more for more variability). This suggests that preferences for tropical wildlife conservation may be well-behaved. Moreover, it highlights the importance of taking into account substitution effects when evaluating scope sensitivity.

To further examine the robustness of preferences we test for order effects. The results for the small good (EA task), tell us that preferences although highly variable, are not significantly affected by presentation order, instead we observe just minor differences at the attribute level. In contrast, for the larger goods, there does seem to be some evidence of order effects, while welfare estimates are not significantly different, a test of overall model equivalence indicates significant differences in taste parameters with respect to the presentation order of the tasks and we observe a number of attribute level differences in both goods. This is particularly significant in the dissimilar treatments which we suggest arises due the increased task complexity in these treatments. Making the choice task too demanding may lead to apparent anomalies.

The finding of high variability in mean WTP is consistent with other studies dealing with tropical biodiversity (such as, Horton et al 2003). In our case, the results may also be influenced in some way by online implementation method¹⁶. The design could also be improved with greater differentiation between attribute levels or greater sample sizes (as per Rollins and Lyke, 1998).

Overall, our results support the observation that evaluation of theoretical consistency of preferences would do well to look beyond simple scope tests by investigating possible substitution effects and other tests with clear expectations. The results suggest some coherency in preferences when taking into account substitution effects in scope tests. And, we find evidence of some procedural invariance for small good and large similar good with respect to ordering effects.

12 REFERENCES

- Ariely, D., Loewenstein, G., and Prelec, D. (2003) "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. *The Quarterly Journal of Economics*, 118(1):73-105.
- Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., & Schuman, H. (1993), "Report of the NOAA panel on Contingent Valuation", *Federal Register*, vol. 58, no. 10, pp. 4602-4614.
- Balmford, A., Gaston, K. J., Blyth, S., James, A., & Kapos, V. (2003) "Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 1046-1050.
- Banerjee, S. and Murphy, J.H. (2005) The Scope Test Revisited, *Applied Economics Letters*, 12: 613-617.
- Bateman, I. J., Carson, R. T., Hanley, N., Day, B. H., Hanemann, M., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R., & Swanson, J. (2002), *Economic Valuation with Stated Preference Techniques A Manual* Edward Elgar, Cheltenham.
- Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D., and Poe, G.L. (2004) On visible choice sets and scope sensitivity. *Journal of Environmental Economics and Management*, 47: 71-93.
- Bateman, I.J., Day, B.H., Jones, A.P., and Jude, S. (2009) Reducing gain-loss asymmetry: A virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management*, 58(1): 106-118.
- Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997) Does Part-Whole Bias Exist? An Experimental Investigation, *The Economic Journal*, 107, 322-332.
- Berrens, R.P., Bohara, A.K., Jenkins-Smith, H.C., Silva, C.L., and Weimer, D., (2004) Information and effort in contingent valuation surveys: application to global climate

¹⁶ Savage and Waldman (2008) find that online implementation results in poorer data quality than mail surveys as online respondents suffer greater fatigue or boredom as they progress through the questionnaire than mail respondents – they suggest that the visual and interactive stimuli of the online survey may in fact place greater cognitive burden on respondents. Further research is required into the effect of online implementation modes on survey responses.

- change using national internet samples. *Journal of Environmental Economics and Management*, 47(2): 331-363.
- Boyle, K.J., desvousges, W.H., Johnson, F.R., dunford, R.W., and Hudon, S.P., (1994) An Investigation of part-whole bias in contingent valuation, *Journal of Environmental Economics and Management*, 27: 64-83.
- Boyle, K.J., Welsh, M.P., and Bishop, R.C. (1993) The Role of Question Order and Respondent Experience In Contingent Valuation Studies, *Journal of Environmental Economics and Management*, 25(1):S80-99..
- Burgess, N. D., D'Amico Hales, J., Ricketts, T. & Dinerstein, E. (2006). Factoring species, non-species values and threats into biodiversity priority-setting across the ecoregions of Africa and its islands. *Biol. Cons.* **127**, 383-401.
- Burgess, N. D., D'Amico Hales, J., Underwood, E., Dinerstein, E., Olson, D., Itoua, I., Schipper, J., Ricketts, T., & Newman, K.. (2004) *Terrestrial ecoregions of Africa and Madagascar: a continental assessment*. Washington DC.: Island Press.
- Carson, R. T. (1998) "Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation", *Ecological Economics*, vol. 24, no. 1, pp. 15-29.
- Carson, R. T. & Mitchell, R. C. (1995), "Sequencing and Nesting in Contingent Valuation Surveys", *Journal of Environmental Economics and Management*, vol. 28, no. 2, pp. 155-173.
- Carson, R. T., Flores, N. E., & Meade, N. F. (2001), "Contingent Valuation: Controversies and Evidence", *Environmental and Resource Economics*, vol. 19, no. 2, pp. 173-210.
- CBD (2006), *Global Biodiversity Outlook 2* Secretariat of the Convention on Biological Diversity, Montreal.
- CBD (2005) *Options for Mobilizing Financial Resources for the Implementation of the Programme of Work by Developing Countries and Countries with Economies in Transition: Note by the Executive Secretary: Ad Hoc Open-Ended Working Group on Protected Areas*, CBD, Montreal, UNEP/CBD/WG-PA/1/3/26 April 2005.
- Chilton, S.M. and Hutchinson, W. G. (2000) A note on the warm glow of giving and scope sensitivity in contingent valuation studies, *Journal of Economic Psychology*, 21(4): 343-349.
- Christie, M., Hanley, N., Warren, J., Murphy, K., Wright, R., & Hyde, T. (2006), "Valuing the diversity of biodiversity", *Ecological Economics*, vol. 58, no. 2, pp. 304-317.
- Fisher, A.C., (1996) The conceptual underpinnings of the contingent valuation method. In: Bjornstad, D.J., Kahn, J.R. (Eds.), *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs*. Edward Elgar, Cheltenham, pp. 19–37.
- Giraud, K.L., Loomis, J.B., and Johnson, R.L. (1999) Internal and external scope in willingness-to-pay estimates for threatened and endangered wildlife, *Journal of Environmental Management*, 56: 221-229.
- Gregory, R. Lichtenstein, S., and Slovic, P. (1993) Valuing environmental resources: A constructive approach. *Journal of Risk Uncertainty*, 7, 177-197.

- Hailu, A., Adamowicz, W. L., & Boxall, P. C. (2000), "Complements, substitutes, budget constraints and valuation - Application of a multi-program environmental valuation method", *Environmental & Resource Economics*, vol. 16, no. 1, pp. 51-68.
- Hoehn, J.P. and Randall, A. (1989) Too Many Proposals Pass the Benefit-Cost Test - Reply, *American Economic Review*, 81(5): 1450-1452
- Hoehn, J.P., (1991) Valuing the Multidimensional Impacts of Environmental Policy: Theory and Methods, *American Journal of Agricultural Economics*, 73(2): 289-299
- Horton, B., Colarullo, G., Bateman, I. J., & Peres, C. A. (2003), "Evaluating non-user willingness to pay for a large-scale conservation programme in Amazonia: a UK/Italian contingent valuation study", *Environmental Conservation*, vol. 30, no. 2, pp. 139-146.
- Jorgensen, B. S. and Syme, G. J. (2000). "Protest Responses and Willingness to Pay: Attitude Toward Paying for Stormwater Pollution Abatement." *Ecological Economics* 33(2):251-265.
- Kahneman, D. & Knetsch, J. L. (1992), "Valuing Public-Goods - the Purchase of Moral Satisfaction", *Journal of Environmental Economics and Management*, vol. 22, no. 1, pp. 57-70.
- Kanninen, B.J. (2000) Optimal Design of Choice Experiments for Nonmarket Valuation In *Proceedings of Workshop on Stated Preference: What do we Know? Where do we go?*, October 12-13, 2000, Washington, 34-58.
- Kling, C. L. & Thomson, C. J. (1996), "The implications of model specification for welfare estimation in nested logit models", *American Journal of Agricultural Economics*, vol. 78, no. 1, p. 103.
- Kontoleon, A & Swanson. T. (2003), "The WTP for Property Rights for the Giant Panda: Can a Charismatic Species be an Instrument for Nature Conservation?" *Land Economics*, 79 (4): 483-499.
- Kramer, R. A. & Mercer, D. E. (1997), "Valuing a Global Environmental Good: US Residents' Willingness to Pay to Protect Tropical Rain Forests", *Land Economics*, vol. 73, no. 2, pp. 196-210.
- Loomis, J. B. & White, D. S. (1996), "Economic benefits of rare and endangered species: summary and meta-analysis", *Ecological Economics*, vol. 18, no. 3, pp. 197-206.
- Loomis, J., Gonzalezcaban, A., & Gregory, R. (1994), "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates", *Land Economics*, vol. 70, no. 4, pp. 499-506.
- Luce, R. D. (1959), *Individual Choice Behaviour: a theoretical analysis* Wiley, New York.
- Marta-Pedroso, C, Freitas, H. and Domingos, T. (2007) Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews, *Ecological Economics*, 62(3-4): 388-398.
- McFadden, D. (1974), "Conditional logit analysis of qualitative choice behaviour," in *Frontiers in Econometrics*, P. Zarembka, ed., Academic Press, New York.
- Neill, H. R. (1995), "The Context for Substitutes in CVM Studies - Some Empirical Observations", *Journal of Environmental Economics and Management*, vol. 29, no. 3, pp. 393-397.

- Ojea, E. and Loureiro, M.L. (2009) Valuation of Wildlife: Revising some additional considerations for scope tests, *Contemporary Economic Policy*, 27(2): 236-250
- ONS, (2008), "First Release: Internet Access 2008 Households and Individuals", 26 August 2008, Office for National Statistics: Newport, UK.
- Plott, Charles R., (1996), "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis," in *Rational Foundations of Economic Behavior*. K. Arrow, E. Colombatto, M. Perleman, & C. Schmidt, eds. London:Macmillan and NY:St. Martin's, pp. 225–50.
- Powe, N. A. and Bateman, I. J. (2004) Investigating Sensitivity to Scope: A Split-Sample Test of Perceived Scheme Realism, *Land Economics*, 80 (2): 258-271.
- Randall, A. and Hoehn, J.P. (1996) Embedding in Market Demand Systems, *Journal of Environmental Economics and Management*, 30: 369-380.
- Rolfe, J., Bennett, J., & Louviere, J. (2000), "Choice modelling and its potential application to tropical rainforest preservation", *Ecological Economics*, vol. 35, no. 2, pp. 289-302.
- Rollins, K. and Lyke, A (1998) The Case for Diminishing Marginal Existence Values, *Journal of Environmental Economics and Management* 36: 324-344.
- Savage, S. and Waldman, D.M. (2008) learning and fatigue during choice experiments: A comparison of online and mail survey modes. *Journal of Applied Econometrics*. 23: 351-371.
- Schulze, W.D., McClelland, G.H., Lazo, J.K. and Rowe, R.D. (1998) Embedding and Calibration in Measuring Non-Use Values, *Resource and Energy Economics*, 20: 163-178.
- Slovic, P., (1995), The construction of preferences, *American Psychologist*, 50: 364-371.
- Spash, C. L. & Hanley, N. (1995), "Preferences, information and biodiversity preservation", *Ecological Economics*, vol. 12, no. 3, pp. 191-208.
- Spash,C (2007) Deliberative monetary valuation (DMV): Issues in combining economic and political processes to value environmental change, *Ecological Economics*, 68: 390-699
- Svedsater, H. (2000), "Contingent valuation of global environmental resources: Test of perfect and regular embedding", *Journal of Economic Psychology*, vol. 21, no. 6, pp. 605-623.
- Veisten, K., Hoehn, H.F., Navrud, S., and Strand, J. (2004). Scope insensitivity in contingent valuation of complex environmental amenities. *Journal of Environmental Management*, 73: 317-331
- White, P. C. L., Gregory, K. W., & Lindley, P. J., & Richards, G. (1997) "Economic Values of Threatened Mammals in Britain: A Case Study of the Otter *Lutra lutra* and the Water Vole *Arvicola terrestris*", *Biological Conservation*, pp: 345-354.