The Effects of the Number of Alternatives in Choice Experiment Questions

Weizhe Weng, Virginia Tech, weizhe11@vt.edu

Mark D. Morrison, Charles Sturt University, mmorrison@csu.edu.au

Kevin J. Boyle, Virginia Tech, kjboyle@vt.edu

Peter C. Boxall, University of Alberta, pboxall@ualberta.ca

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

Abstract: A key issue in the design of a choice experiment is the number of alternatives that subjects choose over, the status quo (SQ) plus one, two, three or more alternatives. This paper compares three choice-experiment designs, SQ+1, which is theoretically incentive compatible and SQ+2 and SQ+3 that do not satisfy conditions of incentive compatibility. Each subject answered only one of the three choice question treatments. We compare response outcomes in terms of estimated model coefficients, a mean-shift analysis of subjects' choice of the SQ alternative, and estimates of willingness to pay. We simultaneously consider matching that can enhance subjects' ability to answer the choice questions and complexity that can impede responses. Matching and complexity are considered in two dimensions: 1) inferred based on features of the choice question designs and 2) stated based on subject responses to survey questions. While the truth is not known, Carson and Groves (2007) theoretical insights on incentive compatibility supports the SQ+1 treatment as a counterfactual to compare the SQ+2 and SQ+3 treatments against. While the analysis provides evidence of matching that can enhance subjects' ability to choose and complexity that can impede choices, the statistical comparison of welfare estimates indicates no difference across treatments.

Key words: Stated preferences, choice experiment, incentive compatibility, status-quo, matching, complexity

The Effects of the Number of Alternatives in Choice Experiment Questions

Weizhe Weng, Mark D. Morrison, Kevin J. Boyle, Peter C. Boxall

1 Introduction

Choice experiments (CEs) are commonly used in to estimate monetary values for characteristics of public and private goods, but there are unresolved considerations in the design of choice questions (Holmes, Adamowicz and Carlsson, 2017). One such consideration is the number of alternatives subjects choose among, the status quo plus one, two, three or more alternatives. Carson and Groves (2007) suggest that when subjects face a single binary choice, one alternative versus the status quo (SQ), they have an incentive to choose their preferred alternative in the context of provision of a public good or service. When the number of alternatives increases, for instance when subjects face two or more alternatives plus the SQ, the subjects' choices may depend on their perceptions of how other people will choose. For example, when a subject believes their preferred alternative is likely to gain the fewest votes, they may choose their second-most preferred alternative, which violates incentive compatibility.

While the Carson and Groves theoretical argument applies to public goods, this issue is also of relevance for private-good applications. There are not incentive considerations regarding how others will respond as with public goods, but the number of alternatives might affect responses for a private good due to increasing number of alternatives increasing the complexity of choices posed to study subjects (Caussade et al. 2005, Boxall, Adamowicz and Moon, 2009, Hensher, 2004, 2006). Complexity is also a consideration for public-good applications.

Thus, the number of alternatives in a question is a core element in the design of choice experiments. Increasing the number of alternatives ostensibly increases the information that is learned from a sample of a limited size. We say ostensibly because there are tradeoffs, as alluded above, that may compromise the information provided by the responses. For example, a potential outcome of increased complexity may be an increased tendency for some subjects to choose the SQ alternative (Samuelson and Zeckhauser, 1988, Swait and Adamowicz, 2001). If the number of alternatives affects responses to choice questions such that the outcome preference estimates vary, this is undesirable for public- and private-good applications.

This paper explores the effects of number of alternatives with regard to preference coefficient estimation, subjects' predilection to choose the status quo alternative, and the implications for value estimation within the context of valuing a public good. We consider three choice question designs; namely status quo plus one alternative (SQ+1), which is theoretically incentive compatible, and status quo plus two and three alternatives (SQ+2 and SQ+3), which do not satisfy theoretical conditions for incentive compatibility. We implemented a split-sample design where each subject answered only one of these question formats.

The application is restoring the Macquarie Marshes in northwest New South Wales, Australia, which is listed as a wetland of international importance under the Ramsar Convention. With the advent of irrigated agriculture in the area, the size and productivity of the marshes has been in decline. A choice experiment was applied in a survey of New South Wales residents to estimate the value they place on wetland restoration. The attributes in the choice questions include the area (size) of the wetlands, frequency of waterbird breeding in the marshes, number of endangered and protected bird species in the marshes, local irrigation employment, and a cost to subjects' households.

We find significant differences between the SQ+1 and SQ+2 treatments, SQ+1 and SQ+3 treatments, and SQ+2 and SQ+3 treatments in terms coefficient estimates when the data are analyzed using a mixed logit model. These test results are confounded by simultaneous information on subjects' choice of the SQ, preference coefficients, standard deviations of preference coefficients and scale factors. To focus on choices of the SQ, we conducted a mean-shift analysis that provides evidence of matching that enhance subjects' ability to choose among alternatives and task complexity that hinders choices. However, when we turn to comparisons of welfare estimates, which avoids multiple confounding effects in the statistical tests and is the policy-relevant outcome of most studies, there are no significant differences in welfare estimates across the three treatments.

2 Previous Research

Designing and implementing a stated-preference survey requires careful attention to detail. A desirable design is framed such that survey subjects have the incentive to provide a true reflection of their preferences when they make choices (Lloyd 2003). Carson and Groves (2007) theoretically demonstrate that when increasing the number of choice alternatives beyond the SQ plus one alternative, subjects may act strategically to avoid a welfare loss in a public-good application. Using lab experiments, Collins and Vossler (2009) find SQ+2 questions have more deviations from induced preferences than SQ+1 questions, which indicates violations of incentive compatibility. Boyle and Özdemir (2009), in a field experiment, find significantly different preference coefficient estimates between SQ+1 and SQ+2 treatments. Volinskiy et al. (2009) also find differences between SQ+1 and SQ+2 formats in terms of coefficient estimates and WTP estimates.

Weighing against incentive compatibility, increasing the number of alternatives subjects choose over can enhance the information learned from a study with a limited sample size. The theory of matching and task complexity can lead to different results between SQ+1 and SQ+>1 treatments. The matching argument suggests that providing multiple, rather than one non-SQ alternative, increases the likelihood subjects are able to select a preferred alternative (Rolfe and Bennett, 2009; Zhang and Adamowicz, 2011, Oehlmann et al. 2017). On the other hand, if the choice context is hard to understand or does not match reality, subjects may misunderstand the choice or distrust the options presented to them (Fischhoff et al. 1999, Louviere, 2006). Such task complexity has at least two potential effects that can lead to subjects being more likely to choose the SQ (Boxall et al. 2009). This can occur if, when facing uncertainty over outcomes, subjects choose to retain what is known, the SQ. Alternatively, when facing a complex choice task; subjects may avoid the decision by choosing the SQ.

The evidence is mixed with regard to the effect on complexity leading subjects to be more likely to choose the SQ. Boxall et al. (2009) found a positive relationship between the number of choice tasks and the tendency to choose at the SQ, while Adamowicz et al. (2011) observed the opposite relationship.[1] Another way to consider uncertainty is to investigate variance; higher levels of uncertainty may lead to larger variance in estimation. DeShazo and Fermo (2002) find that the relationship between the number of alternatives and the variance of utility exhibits a quadratic relationship. This result provides evidence that uncertainty, reflected in variance of utility, will first decrease, perhaps due to enhanced preference matching, and then increase due to task complexity.

[1] See also: Boxall (2009), Adamowicz et al. (2011), Rolfe and Windle (2012), Meyerhoff et al. (2015), and Oehlmann et al. (2017).

Thus, the empirical evidence on the impact of posing a choice question with two or more alternatives in addition to the SQ is mixed. The existing literature suggest that there is not a single uniform outcome observed and general insights will be developed through a number of systematic investigations with varying number of choice alternatives in experimental treatments in the context of different applications and samples of subjects.

In this study, we investigate differences between the SQ+1 and SQ+2 treatments, SQ+1 and SQ+3 treatments, and SQ+2 and SQ+3 treatments, in terms of subjects' choice of the SQ, coefficient estimates and welfare estimates of willingness to pay for stylized policy scenarios. We consider choice complexity in two dimensions, inferred complexity (Boxall et al. 2009) and subjects' stated complexity. Inferred complexity is based on the number of attribute level changes in choice questions and the number of choice tasks. Stated complexity is based on subjects' answers to complexity questions. We investigate how stated and inferred complexity influence subjects' decisions to select the SQ.

3 Study Area and Survey Design

The application is valuing restoration of the Macquarie Marshes, which is an ephemeral wetland on the Macquarie River in northwest New South Wales, Australia. A nature reserve, contained in the Marshes, is listed as a wetland of international importance under the Ramsar Convention. The Marshes have a number of significant environmental characteristics. They provide an important habitat for waterbird habitat and breeding, act as a filter that improves downstream water quality, and provide high-quality feed for livestock and other services as well.

The Macquarie Marshes were originally the largest wetlands in NSW with an area of approximately 2200 km^2. However, due to the use of water for irrigation and lengthy droughts, the health of the marshes and area of the marshes had declined dramatically. The area of the Marshes has fallen from to about 500 km^2, the frequency of waterbird breeding has declined from occurring almost every year on average to every eight years, and the number of endangered and protected bird species using the Marshes as habitat has fallen from 31 to six species.

3.1 Survey Design

Previous studies have estimated values for restoring the Macquarie Marshes (Morrison et al., 1999 and 2002). We use these previous studies to inform the design of the current survey and the estimation results to inform the design of the choice experiment.

The survey was designed to be administered via the internet. The survey contained a description of the Macquarie Marshes and the related environmental problems, a description of possible options for marsh restoration, choice questions, and socio-demographic, opinion and attitudinal questions. Subjects were told that the size of the marshes had declined due to the use of water for irrigation and climate change. To accomplish improvements in the marshes subjects were also told that the government would purchase water rights from farmers using the existing water trading market. Subjects would pay for these purchases through a one-time increase in water rates.

Subjects were told that one alternative is to continue the current situation (SQ). Subjects were presented choice questions with one, two or three alternatives to increase the amount of water allocated to the Macquarie Marshes, which would decrease irrigation employment while increasing the size and health of the marshes. An example choice question (SQ+2) is shown in Figure 1. The attributes in the choice-questions included wetland area restored, waterbird breeding frequency, the number of endangered and protected waterbird species using the Marshes as habitat, irrigation-related employment, and the cost to subjects' household. The levels for each attribute are shown in Table 1.[2] Subjects were asked to answer eight choice questions for each of the SQ+1, SQ+2 and SQ+3 treatments.[3]

4 Econometric Modelling

There are four elements in the analysis – model estimation, computation of welfare estimates, mean-shift analysis, and hypotheses investigated. The mean shift analysis considers whether increasing the number of alternatives leads subjects to be more likely to choose the SQ.

4.1 Model estimation

Let U_{nit} denote the utility of respondent n for alternative i in choice situation t:

$$U_{nit} = V_{nit} + \varepsilon_{nit} \tag{1}$$

where V_{nit} is the systematic component that varies over people, ε_{nit} is the random error that represents the unobserved component of utility and is assumed to be *iid* extreme value. We assume the systematic component of utility is linear in the choice-question attributes as:

$$V_{nit} = \beta_{ASC}(ASC) + \beta_1(cost) + \beta_2(emp) + \beta_3(wet) + \beta_4(bird) + \beta_5(end) \tag{2}$$

where the *ASC* equals 1 for the SQ alternative and 0 otherwise, the attribute variables are measured in the units shown in Table 1, the β's are coefficients to be estimated.

[2] A Db-efficient attribute design was used. This is a Bayesian approach where the analyst applies a set of prior parameters to infer the attribute level combinations that will minimize the elements within the expected asymptotic variance-covariance matrix. Parameter estimates from Morrison et al. (1999, 2002) were used to inform this process. Rather than assuming precise knowledge of the population parameter estimates, Db-efficient designs utilize distributions of likely parameter estimates in the design process.
[3] We recognize that a sequence of eight questions may violate the incentive compatibility properties of a single binary question. Subjects were to told answer each choice question as it was the only alternative available to choose over and they could not go back and change answers to choice questions they had already answered.

Contemporary analyses of choice-experiment data allow for preference heterogeneity across respondents and correlated responses across choice questions (McFadden & Train, 2000). There are several different modeling variants to address these considerations. Mixed logit models are estimated that allows all sources of correlation, including correlation induced by scale heterogeneity, in the model (Hess and Train, 2017). All coefficient estimates are assumed to be normally distributed, except cost that is assumed to be fixed to facilitate calculation of welfare estimates.

4.2 Welfare Estimation

Value estimates are computed using the mixed logit coeeficent estimates from equation (2) as follows:

$$WTP = (\sum_{a=1}^{A} \hat{\beta}_a * (x_a^{imp} - x_a^{SQ}))/\hat{\beta}_c \qquad (3)$$

where the $\hat{\beta}_a$ are estimated coefficents for the nonprice attributes and $\hat{\beta}_c$ is the estimated coefficent for the price attribute, and *imp* denotes an improved attribute level and SQ denotes a staus quo attribute level.

One policy scenarios is considered. We compute the willingness to pay for restoring environmental attributes to their historical highest levels with no change in employment, WTP_{HIGH}. In this case, the wetland area of the Macquarie marshes increases from 500 to 2200 km^2, waterbird breeding frequency increases from every eight years to every year, and number of endangered species increases from six to 31.

4.3 Mean-shift analysis

The sign and significance of the ASC coefficients reveals whether the subjects, as a whole, tend to choose the SQ or not. Additional analysis is needed to investigate if increasing task complexity leads subjects to be more likely to choose the SQ.

In the mixed logit model, we assume β_{ASC} is randomly distributed $N(\mu_{SQ}, \sigma_{SQ}^2)$. In other words, respondents are heterogeneous in their preference for the SQ and each subject can have a unique ASC coefficient. Another way to investigate preference heterogeneity for the status quo is to have a random ASC coefficient and interact the ASC variable with subject and study design characteristics. This is the mean-shift analysis (Boxall et al., 2009).

Choice-question responses are not pooled across treatments (SQ+1, SQ+2 and SQ+3) and equation (2) is expanded to include respondent characteristics and study treatment features as regressors to investigate if task complexity influences choices of the SQ as follows:

$$V_{nit} = \beta_{ASC}(ASC) + \beta_1(cost) + \beta_2(emp) + \beta_3(wet) + \beta_4(bird) + \beta_5(end) + \gamma_{ASC} * ASC * \boldsymbol{Z}_{t}. \quad (4)$$

The new component of the equation is a vector, $\mathbf{Z}_t$, which consists of respondent demographic and opinion regressors, and indicators of task complexity regressors. The $\gamma_{ASC}'s$ are coefficients to be estimated that represent the 'shift' in the mean of the SQ coefficient (β_{ASC}) given that each variable in Z_t is multiplied by *ASC*. Equation (4) is again estimated as a mixed logit model.

The demographic and opinion characteristic regressors are:

Farm – a member of family is associated with farming,

Memb – member of an organization with environmental conservation focus, *Age* –age,

Edu –has post-secondary education,

Inc –household income,

OpPurch – irrigation water should be purchased from farmers,

OpBias – information in the survey biased in favor of the wetland,

OpPay – payment to improve environmental is a good idea,

OpWork – purchasing irrigation water from the farmers would work, and

OpTrust – trust the increase in water rates will be one-off.

The choice-complexity regressors include:

#S – level for an attribute changes once across alternatives within a choice question,[4]

#D – level for an attribute changes twice across alternatives within a choice question,[5]

#M – level for an attribute changes more than two times across alternatives within a choice question),[6]

Task# – choice set number in the sequence of the total number of tasks,

---

[4] *#S* ranges from 8 to 40 for the SQ+1 treatment, where one attribute level must change in each of the 8 choice questions. The maximum is defined when all attributes change for all choice questions: (#of attributes)*(# of choice sets). *#S* can be zero for the SQ+2 and SQ+3 treatments, when *#D* or *#M* are nonzero.

[5] *#D* ranges from 0 to 40 for the SQ+2 and SQ+3 treatments, where the minimum can be zero when *#S* or *#M* are nonzero. The maximum is defined when all attributes change twice for all choice questions: (#of attributes)*(# of choice questions). *#D* is zero for the SQ+1 because multiple attribute level changes are not possible for this question format.

[6] *#M* ranges from 0 to 40 for the SQ+3 treatment, where the minimum can be zero when *#S* or *#D* are nonzero. The maximum is defined when all attributes change more than two times for all choice questions: (#of attributes)*(# of choice questions). *#M* is zero for the SQ+1 and SQ+2 treatments because more than two attribute level changes are not possible for this question format.

InfoUnd – low understanding of the information in the survey,

InfoMore – need more information than provided,

InfoConf – information in the choice set confusing, and

AnswDiff – picking an alternative is hard.

S#, D#, M# and *Task#* are objective descriptors of the choice questions and are indicators of inferred complexity.[7] *InfoUnd, InforMore, InfoBias, InfoConf* and *AnswDiff* are subjects' responses to survey questions and are the indicators of stated complexity. The coefficient estimates for these two sets of variables are of primary interest in the mean-shift analysis.

Intuition for interpreting the γ_{ASC} coefficients for task complexity might be as follows. Among the inferred complexity variables, if the number of attribute level changes (*#S, #D* and *#M*) facilitates matching, we would expect the associated γ coefficient to be negative, subjects be less likely to choose the SQ alternative. The opposite is expected if the number of attribute changes increases complexity. Alternatively, it could be that this feature is neutral; it does not influence how subjects answer the choice questions, and is insignificant. Choice complexity is indicated by a positive and significant coefficient on *Task*; complexity increases as additional choice questions are answered. With the internet implementation, subjects could not change their answers to previous choice questions.

For stated-complexity, low understanding (*InfoUnd*), needing more information (*InfoMore*), confusing information (*InfoConf*) and difficulties in pick choices (*AnswDiff*) reflect

[7] Others have considered choice complexity in terms of entropy, which is a cumulative measure of the elements of complexity in a set of choice questions (Zhang and Adamowicz 2011; Oehlmann et al. 2017; Swait and Adamowicz 2011). We use a more disaggregated measures of complexity to investigate the effects of different elements of complexity on responses following Boxall et al. (2009).

complexity, while the alternative levels of these variables might enable matching. For these variables, positive and significant coefficients indicate that complexity leads subjects to be more likely to choose the SQ alternative, negative and significant supports matching, and insignificance indicates no effect.

4.4 Hypotheses Investigated

When conducting hypothesis tests of estimates using mixed logit estimation, differences in preference heterogenity and scale cannot be disentangled (Hess and Rose, 2012). Further, tests of model estimates across experimental treaments include information on preferences and the distribution of preferences. It is not possible to do tests that solely indicate differences in preferences nor solely difference in the distribution of prteference.

While the primary interest is wheteher preference coefficient estimates are constant across treatments, the complications disucced above preclude conducting such a test. We can simply identify if there is a statisticlly significant difference in estimation results across treatements. We therefore conduct likelihood-ratio tests, following Swait and Louviere (1993), to investigate whether preference estimates and estimates of their distributions are collectively different across treatment; this captures the combined effects of the *ASC*'s, attribute coefficients and coefficent standard deviations.[8] Thus, the following hypotheses is formulated:

$$H_0\colon \beta_{(SQ+1)} = \beta_{(SQ+2)}\ vs.\, H_a\colon \beta_{(SQ+1)} \neq \beta_{(SQ+2)} \tag{5a}$$

$$H_0\colon \beta_{(SQ+1)} = \beta_{(SQ+3)}\ vs.\, H_a\colon \beta_{(SQ+1)} \neq \beta_{(SQ+3)} \tag{5b}$$

$$H_0\colon \beta_{(SQ+2)} = \beta_{(SQ+3)}\ vs.\ H_a\colon \beta_{(SQ+2)} \neq \beta_{(SQ+3)} \tag{5c}$$

[8] The likelihood-ratio test statistic is λ=-2[L(pooled(i,j)-(L(SQ+i) +L(SQ+j))] ∀ $i \neq j$.

The first two hypotheses ask if the estimation results for the incentive-compatible question format (SQ+1) are the same as the non-incentive-compatible question formats (SQ+2 and SQ+3). While the third hypothesis asks if the results from the two non-incentive-compatible formats are statistically equivalent. If the null hypotheses (5a, 5b or 5c) cannot be rejected, this is evidence of equivalence, but not confirmatory evidence because of the confounding issues discussed above. These tests are conducted using the mixed logit coefficient estimates based on equation (2).

To investigate the equivalence of the estimation results we turn to comparisons of willingness to pay as defined in equation (3). The ration of attribute coefficients to the cost coefficient in this equation cancels out the effect of the scale parameter and the coefficient standard deviations are removed from these tests. It is the case that the welfare calculation is often the most important estimate for environmental policy applications, be it a marginal part worth for a single attribute or an aggregate value based on simultaneous changes in multiple attributes. The hypotheses tested are.

$$H_0\colon WTP_{(SQ+1)} = WTP_{(SQ+2)}\ vs.\ H_a\colon WTP_{(SQ+1)} > WTP_{(SQ+2)} \tag{6a}$$

$$H_0\colon WTP_{(SQ+1)} = WTP_{(SQ+3)}\ vs.\ H_a\colon WTP_{(SQ+1)} > WTP_{(SQ+3)} \tag{6b}$$

$$H_0\colon WTP_{(SQ+2)} = WTP_{(SQ+3)}\ vs. H_a\colon WTP_{(SQ+2)} > WTP_{(SQ+3)} \tag{6c}$$

and

$$H_0\colon WTP_{(SQ+1)} = WTP_{(SQ+2)}\ vs.\ H_a\colon WTP_{(SQ+1)} < WTP_{(SQ+2)} \tag{7a}$$

$$H_0\colon WTP_{(SQ+1)} = WTP_{(SQ+3)}\ vs. H_a\colon WTP_{(SQ+1)} < WTP_{(SQ+3)} \tag{7b}$$

$$H_0: WTP_{(SQ+2)} = WTP_{(SQ+3)}\ vs.\ H_a: WTP_{(SQ+2)} < WTP_{(SQ+3)} \tag{7c}$$

Convolution tests are used to investigate these hypotheses and two sets of tests are conducted since convolutions provides one-sided test reulsts (Poe, Giraud and Loomis 2005). Again the first two hypotheses in each set test the incentive compatible treatment (SQ+1) against each of the treatments that are not incentinve comaptible (SQ+2 and SQ+3). Failure to reject the null hypotheses (6 and 7) is evidence that the treatments provide identical information on WTP and matching or complexity has not affected welfare estimates across treeatments.

5 Empirial Results

Subjects were randomly recruited from an online panel of NSW residents provided by *Research Now*. Subjects were directed to a website hosted by the Institute for Transport Studies at Sydney University. Each subject was randomly assigned to one of the treatment groups, SQ+1, SQ+2 or SQ+3.

A total of 1,827 respondents answered all eight choice questions. The sample sizes for the treatment are N_{SQ+1} = 609, N_{SQ+2}=622 and N_{SQ+3}=596. The response rates are statistically equivalent across treatments (χ^2=0.814, p=0.66).

Summary data on respondents' characteristics by treatment are presented in Table 2 and some minor differences are noted. SQ+1 respondents are more likely to have a family member associated with farming than SQ+2 respondents, but the absolute difference is not large (15% v. 12%). The SQ+2 respondents are older population on average that the SQ+1 and SQ+3 respondents, but the absolute difference is only two years (39 versus 41 years of age). Respondents in the SQ+2 treatment were more likely to say the information was biased than the SQ+1 respondents, 37% versus 31%. Controlling for these variables do not affect attribute coefficient estimates. All the same, we do control for these respondent characteristics in the mean-shift analysis.

Differences are observed in response to the inferred complexity questions (Table 2). SQ+1 respondents are less likely to indicate the choice-questions information was confusing (*InfoConf* - 11% versus 15% and 16%) and this is also true for difficulty picking an alternative (*AnswDiff* - 12% versus 17% and 16%), which are indications of increasing choice complexity when respondents were presented with more alternatives to choose among. We investigate if these conditions lead respondents to be more likely to choose the SQ in the mean shift analysis.

5.1 Model estimates

The mixed logit estimates for equation (2) are presented in Table 3 by treatment. All attribute coefficients have the expected signs and are significant at the 1% level. Respondents prefer more more employment, a larger wetland area, more frequent waterbird breeding, more endangered species, and the cost coefficient is negative. The standard deviations of distributed coefficients are also all significant at the 1% level.

Note also, while it has been argued that adding alternatives can enhance the information from a study as subjects choose over more combinations of attribute combinations and levels in choice questions, the SQ+1 model has the lowest goodness of fit meanures in terms of the log likelihoob, AIC and BIC measures. These goodness of fit measures increase as the number of alterantives presented to subjects increases. This is suggestive evidence of increasing task complexity.

A surprising result that differs from empirical results in previous majority of the literatue is that the ASC coefficients are all significant and are negative. This indicates that factors other the the design attributes make respondents likely to not choose the SQ, which is the oposite of what has been found in previous studies (e.g., Boxall et al. 2009; Zhang and Adamowicz et al. 2011; Oehlmann et al. 2017).

Among three treatments, we find that respondents are most likely in absolte terms to choose the SQ in the SQ+1 treatment, followed by the SQ+3 and SQ+2 treatments. The SQ was chosen in 47% of choice taskes for SQ+1, 37% for SQ+3, then reduced to 28% for SQ+2. A χ^2-test for equal frequency of choosing the SQ in the three treatments is rejected (χ^2= 563.24, p<0.001). If subjects are less likely to choose the SQ when presented with more altrernatives to choose over, this is suggestive evidence of matching.

5.2 Willingness to pay estimates

WTP_{HIGH} is computed for restoring environmental attributes to historical levels with no change in employment. In this case, the wetland area increases from 500 to 2200 km^2, waterbird-breeding frequency increases from every eight years to every year, and the number of endangered species increases from six to 31:

$$WTP_{HIGH} = \frac{\beta_{wet}*(2{,}200-500)+\beta_{bird}*(1-8)+\beta_{end}*(31-6)}{|\beta_{cost}|}. \tag{8}$$

These welfare estimates are presented for expository purposes and are not linked to a specific policy proposal. WTP estimates by treatment are reported in in Table 4, with no observable pattern of results that provides evidence of matching or complexity. Bootstrapping using 500 draws is used to compute 90% confidence intervals.

5.3 Mean-shift analysis

Estimation results for equation (4) are presented in Table 5.[9] Similar with the original mixed logit model (equation 2, Table 3), all attribute coefficients have the expected signs and are significant at the 1 percent level. Comparison of attribute coefficient estimates, by treatment, in Tables 3 and 5 reveals nearly identical results, suggesting that the inclusion/exclusion of respondent demographic characteristics does not affect coefficient estimates of the key policy variables. The magnitudes of the ASC coefficient estimates change from Table 3 to Table 5; this is expected given the addition of interaction terms with the ASC in equation (4) (Table 5). Standard deviations of distributed coefficients are significant at 1% level (see Appendix Table).

Turning to the influence of context variables, we find that inferred complexity affects respondents' choices of the SQ. For the SQ+1 treatment, increasing number of single attribute changes results in respondents being less likely to choose the SQ, which suggests these changes may have facilitated matching. For the SQ+3 treatment, increasing number of double and multiple attribute changes leads to respondents tending to be more likely to choose the SQ, suggesting a choice-complexity effect. This may be supportive evidence for the quadratic relationship observed by DeShazo and Fermo (2002).

The coefficients on task sequence are positive and significant in all three treatments. This result indicates that as additional questions are asked subjects become more likely to choose the SQ. This may be from confusion across questions/alternatives or fatigue from answering multiple, choice questions.

[9] With a large number of right-hand side variables, multicollinearity becomes a concern. Thus, we have calculated the variance inflation factors (VIF) of each variable. In all cases, the calculated VIF is less than 10 (except for the VIFs for interaction terms with ASC), so we keep all the variables on the right-hand side.

Stated complexity had less effect on respondents' choices of the SQ; the only significant coefficients are for *InfoUnd* for treatments SQ+1 and SQ+2. The positive and significant coefficients of indicates low understanding leads respondents to be more likely to choose the SQ. Thus, limited understanding may be associated with increased task complexity, but it is surprising that this effect is not significant for the SQ+3 treatment.

5.4 Hypothesis Test Results

The likelihood ratio test results indicate that the null hypotheses (5) of equivalence of coefficent estimates between the SQ+1 and SQ+2, SQ+1 and SQ+3, and SQ+2 and SQ+3 can be rejected. This is evidence that the number of alternatives in a choice question affects testimation results, but the results of these tests are confound over coefficent estimates, their standard deviations and scale paramteters. The difference in preferences may come from the differences in the SQ effects, attribute preferences, or the confounding effects of standard deviation and scale emasures of heterogeneity. At a minimum, the mean shift analysis results suggests there are matching and complexity effects, and that these effects may not be constant across treatments with differring numbers of alternatives.

When we move to the comparisons of the willingness to pay (hypotheses 6 and 7), there is no difference in the WTP estimates in for either scenario across the three treatments. Thus the differences identified in the test of coefficent estmates does not carry over to the policy-reelvant information, welfare estimates.

6 Conclusions

This paper compared three choice-experiment designs, SQ+1, which is theoretically incentive compatible and SQ+2 and SQ+3 that do not satisfy conditions of incentive compatibility. While the truth is not known, Carson and Groves (2007) theoretical insights on incentive compatibility supports the SQ+1 treatment as a counterfactual to compare the SQ+2 and SQ+3 treatments against. While the analysis provides evidence of matching that can enhance subjects' ability to choose and complexity that can impede choices, the statistical comparison of welfare estimates indicates no difference across treatments.

The practical implications of these results are three fold. First, we observe mixed results across treatments, which is consistent with the previous literature; some of our results support the previous findings and some do not. Second, from an outcome perspective, these differences may not be of consequence for policy applications of study results as estimates of WTP are statistically invariant across the three treatments, indicating that choice questions with a SQ plus one, two or three alternatives provide similar economic insights. Third, the carrying insights on matching and complexity do suggest that there remains room where careful survey design may enhance the quality of empirical outcomes, e.g., ensuring the information provided in the survey is clear and understood by subjects.

This continues to be an area of investigation that warrants further insights.

References

Adamowicz, Wiktor, Diane Dupont, Alan Krupnick, and Jing Zhang. 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. Journal of Environmental Economics and Management 61 (2):213-226.

Adamwicz, V, P Boxall, M Williams, and J Louviere. 1998. Stated preference approaches to measuring passive use values. American Journal of Agricultural Economics 80:64-75.

Arentze, Theo, Aloys Borgers, Harry Timmermans, and Romano DelMistro. 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E: Logistics and Transportation Review 39 (3):229-244.

Bonnichsen, Ole, and Jacob Ladenburg. 2015. Reducing status quo bias in choice experiments. Nordic Journal of Health Economics 3 (1):pp. 47-67.

Boxall, Peter, Wiktor L Adamowicz, and Amanda Moon. 2009. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. Australian journal of agricultural and resource economics 53 (4):503-519.

Boyle, Kevin J., and Semra Özdemir. 2009. Convergent Validity of Attribute-Based, Choice Questions in Stated-Preference Studies. Environmental and Resource Economics 42 (2):247-264.

Breffle, William S, and Robert D Rowe. 2002. Comparing choice question formats for evaluating natural resource tradeoffs. Land Economics 78 (2):298-314.

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

Carson, Richard T, Theodore Groves, and John A List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. Journal of the Association of Environmental and Resource Economists 1 (1/2):171-207.

Caussade, Sebastian, Juan de Dios Ortúzar, Luis I Rizzi, and David A Hensher. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation research part B: Methodological 39 (7):621-640.

Collins, Jill P, and Christian A Vossler. 2009. Incentive compatibility tests of choice experiment value elicitation questions. Journal of Environmental Economics and Management 58 (2):226-235.

Day, Brett, Ian J Bateman, Richard T Carson, Diane Dupont, Jordan J Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63 (1):73-91.

DeShazo, JR, and German Fermo. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 44 (1):123-143.

Fischhoff, Baruch, Ned Welch, Shane Frederick, Jeff Dominitz, and Timothy L McDaniels. 1999. Construal processes in preference assessment. In Elicitation of Preferences: Springer, 139-170.

Hensher, David A. 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. Journal of Transport Economics and Policy (JTEP) 38 (3):425-446.

———. 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. Environmental and Resource Economics 34 (1):7-44.

Hess, S., & Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models?. Transportation, 39(6), 1225-1239.

Hess, S., & Train, K. (2017). Correlation and scale in mixed logit models. Journal of Choice Modelling, 23, 1-8.

Holmes, Thomas P, and Kevin J Boyle. 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics 81 (1):114-126.

List, John A. 2007. Using experimental methods in environmental and resource economics: Edward Elgar Publishing.

Lloyd, Andrew J. 2003. Threats to the estimation of benefit: are preference elicitation methods accurate? Health economics 12 (5):393-402.

Louviere, Jordan J. 2006. What you don’t know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. Environmental and Resource Economics 34 (1):173-188.

McFadden, Daniel, and Kenneth Train. 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics:447-470.

Meyerhoff, Jürgen, and Ulf Liebe. 2009. Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity. Land Economics 85 (3):515-528.

Meyerhoff, Jürgen, Malte Oehlmann, and Priska Weller. 2015. The Influence of Design Dimensions on Stated Choices in an Environmental Context. Environmental and Resource Economics 61 (3):385-407.

Morrison, Mark, Jeff Bennett, and Russell Blamey. 1999. Valuing improved wetland quality using choice modeling. Water Resources Research 35 (9):2805-2814.

Morrison, Mark, Jeff Bennett, Russell Blamey, and Jordan Louviere. 2002. Choice modeling and tests of benefit transfer. American Journal of Agricultural Economics 84 (1):161-170.

Oehlmann, Malte, Jürgen Meyerhoff, Petr Mariel, and Priska Weller. 2017. Uncovering context-induced status quo effects in choice experiments. Journal of Environmental Economics and Management 81:59-73.

Poe, Gregory L, Kelly L Giraud, and John B Loomis. 2001. Simple computational methods for measuring the difference of empirical distributions: application to internal and external scope tests in contingent valuation. staff paper 5.

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

Rolfe, John, and Jill Windle. 2012. Restricted versus Unrestricted Choice in Labeled Choice Experiments: Exploring the Trade‐Offs of Expanding Choice Dimensions. Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie 60 (1):53-70.

Samuelson, William, and Richard Zeckhauser. 1988. Status quo bias in decision making. Journal of Risk and Uncertainty 1 (1):7-59.

Scarpa, Riccardo, Silvia Ferrini, and Kenneth Willis. 2005. Performance of error component models for status-quo effects in choice experiments. In Applications of simulation methods in environmental and resource economics: Springer, 247-273.

Swait, Joffre, and Wiktor Adamowicz. 2001. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. Journal of Consumer Research 28 (1):135-148.

Swait, Joffre, and Jordan Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of marketing research:305-314.

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

Volinskiy, Dmitriy, Wiktor L Adamowicz, Michele Veeman, and Lorie Srivastava. 2009. Does Choice Context Affect the Results from Incentive‐Compatible Experiments? The Case of Non‐GM and Country‐of‐Origin Premia in Canola Oil. Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie 57 (2):205-221.

Zhang, Jing, and Wiktor L Adamowicz. 2011. Unraveling the choice format effect: A context-dependent random utility model. Land Economics 87 (4):730-743.

Table 1. Attribute and attribute levels

Attributes	Status Quo	Attribute Levels
cost-water rates (one-off increase in AUD)	no change	\$20, \$50, \$75, \$100, \$125, \$150, \$200, \$250
emp-irrigation related employment	4400	4200, 4000, 3800 jobs
wet-Wetlands area	500	700, 900, 1100 km^2
bird-waterbirds breeding	every 8 years	every 6, 4, 2 years
end-endangered and protected bird species present	6 species	12, 18, 25 species

Table 2. Socio-demographic characteristics of respondents and attitudinal characteristics of the samples (significant differences between treatments at the 10% level denoted by superscripts)[a]

	SQ+1	SQ+2	SQ+3
Socio-demographic characteristics			
Farm	15%[SQ+2]	12%[SQ+1]	13%
Memb	7%	6%	6%
Age	39[SQ+2]	41[SQ+1,SQ+3]	39[SQ+2]
Edu	39%	40%	37%
Inc	67444	67669	66175
Attitudinal characteristics			
OpPurch	57%	58%	56%
OpBias	31%[SQ+2]	37%[SQ+1]	32%
OpPay	57%	55% [SQ+3]	60% [SQ+2]
OpWork	55%	57%	56%
OpTrust	6%	6%	7%
Stated complexity			
InfoUnd	1%	2%	2%
InfoMore	18%	19%	21%
InfoConf	11%[SQ+2,SQ+3]	15%[SQ+1]	16%[SQ+1]
AnswDiff	12%[SQ+2,SQ+3]	17%[SQ+1]	16%[SQ+1]

[a]. Superscripts denote statistically significant differences at the 10% level, e.g., the SQ+2 superscript on the SQ+1 age statistics indicates that the statistics for these two treatments are significantly different at 10% level.

Table 3. Mixed logit estimates

	Coefficient Estimates		
	SQ+1	SQ+2	SQ+3
ASC_{SQ}	-1.855***[a]	-2.290***	-2.066***
	(0.237)	(0.175)	(0.229)
cost	-0.023***	-0.014***	-0.013***
	(0.001)	(0.001)	(0.000)
emp	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
wet	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
bird	-0.184***	-0.124***	-0.124***
	(0.034)	(0.014)	(0.016)
end	0.907***	0.061***	0.053***
	(0.011)	(0.005)	(0.006)
ASC_{SQ}-sd[b]	2.959***	2.915***	4.726***
	(0.316)	(0.231)	(0.374)
emp-sd	0.003***	0.002***	0.0001***
	(0.000)	(0.000)	(0.000)
wet-sd	0.004***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
bird-sd	0.384***	0.169***	0.156***
	(0.042)	(0.022)	(0.020)
end-sd	0.131***	0.071***	0.093***
	(0.013)	(0.006)	(0.007)
Log likelihood	-2279.589	-3724.451	-4197.455
AIC	4601.177	7490.901	8436.91
BIC	4752.05	7650.732	8601.886

[a] ***p<0.01

[b] sd denotes standard deviations of normal distributed coefficients

Table 4. Willingness to pay estimates (AUD$s)

	WTP_{HIGH}
SQ+1	$288***[a]
	($212 , $365)[b]
SQ+2	$263***
	($197 , $329)
SQ+3	$308***
	($239 , $378)

[a]***p<0.01

[b]90% confidence interval in parentheses

Table 5. Mean-shift estimates

	Coefficient Estimates		
	SQ+1	SQ+2	SQ+3
ASC and attributes			
ASC_{SQ}	0.768	0.142	1.374
	(0.594)	(1.419)	(1.517)
cost	-0.023***[a]	-0.014***	-0.013***
	(0.001)	(0.001)	(0.000)
emp	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
wet	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
bird	-0.113***	-0.127***	-0.140***
	(0.038)	(0.015)	(0.016)
end	0.715***	0.064***	0.060***
	(0.012)	(0.005)	(0.006)
Socio-demographic characteristics			
ASC*Farm	-0.611**[b]	-0.110	-0.715
	(0.286)	(0.460)	(0.476)
ASC*Memb	-0.953**	0.020	-2.065***
	(0.476)	(0.490)	(0.600)
ASC*Age	-0.024**	-0.031***	-0.062***
	(0.011)	(0.009)	(0.018)
ASC*Edu	-0.089	-0.235	0.034
	(0.254)	(0.292)	(0.458)
ASC*Inc	-0.000003	-0.000002	-0.00001**
	(0.000)	(0.000)	(0.000)
Inferred complexity			
ASC*S	-0.264***	-0.028	0.212
	(0.074)	(0.271)	(0.249)
ASC*D	-	-0.052	0.567**
		(0.271)	(0.254)
ASC*M	-	-	0.454*
			(0.265)
ASC*Task	0.053**	0.063***	0.100***
	(0.022)	(0.022)	(0.028)
Stated complexity			
ASC*InfoUnd	1.606***	1.263*[c]	0.951
	(0.615)	(0.685)	(0.682)
ASC*InfoMore	-0.165	0.077	-0.503
	(0.362)	(0.302)	(0.424)
ASC*InfoConf	-0.461	-0.022	-0.114
	(0.427)	(0.373)	(0.643)
ASC*AnswDiff	0.294	-0.137	-0.137
	(0.343)	(0.372)	(0.580)
Attitudinal characteristics			
ASC*OpPurch	0.093	-0.102	-2.546***
	(0.310)	(0.321)	(0.424)
ASC*OpBias	0.836***	0.868***	1.087***
	(0.269)	(0.285)	(0.389)
ASC*OpPay	-1.040***	-1.959***	-2.029***
	(0.276)	(0.308)	(0.418)
ASC*OpWork	-1.217***	-0.069	-1.214*
	(0.302)	(0.326)	(0.471)
ASC*OpTrust	-0.258	-0.321	0.640
	(0.466)	(0.473)	(0.648)
Log likelihood	-2230.261	-3698.907	-4142.457
AIC	4534.521	7473.813	8362.914
BIC	4800.344	7763.031	8669.297

[a]***p<0.01, [b]**p<0.05, [c]*p<0.1

Table 6. Hypothesis results

	SQ+1 vs. SQ+2	SQ+1 vs. SQ+3	SQ+2 vs. SQ+3
		Hypothesis (5)	
Mixed logit model	156.29[a]***[b]	266.79***	151.11***
		Hypothesis (6)	
WTP_{HIGH}	0.522[c]	0.751	0.751
		Hypothesis (7)	
WTP_{HIGH}	0.479[d]	0.249	0.250

[a] Likelihood ratio chi-square test statistic
[b] ***p<0.01
[c] p-value for one-sided convolution test
[d] p-value for one-sided convolution test

Appendix Table. Standard deviation of mean-shift analysis

	Standard Deviation Estimates		
	SQ+1	SQ+2	SQ+3
ASC_{SQ}-sd[a]	2.943***	2.621***	4.886***
	(0.322)	(0.198)	(0.359)
emp-sd	0.004***	0.002***	0.0001***
	(0.000)	(0.000)	(0.000)
wet-sd	0.004***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
bird-sd	0.412***	0.159***	0.162***
	(0.045)	(0.021)	(0.022)
end-sd	0.115***	0.068***	0.086***
	(0.013)	(0.006)	(0.009)

[a] sd denotes standard deviations of normal distributed coefficients

Figure 1. Example choice question

Suppose options 1, 2 and 3 are the **ONLY** ones available, realistically which one would you choose?

Outcome	Option 1: Continue current situation	Option 2: Increase water to Macquarie Marshes	Option 3: Increase water to Macquarie Marshes
Your water rates (one-off increase)	no change	$20 increase	$50 increase
Irrigation related employment	4400 jobs	4350 jobs	4350 jobs
Wetlands area	500 km^2	650 km^2	1000 km^2
Waterbirds breeding	every 8 years	every 3 years	every year
Endangered and protected bird species present	6 species	25 species	15 species
I would choose	❒	❒	❒