Do Choice Experiments Generate Reliable Willingness to Pay Estimates?

Theory and Experimental Evidence

Katherine Silz Carson

Department of Economics and Geosciences, United States Air Force Academy

Susan M. Chilton

Newcastle University Business School – Economics

W. George Hutchinson

Gibson Institute for Land, Food and Environment, School of Biological Sciences, Queen's University, Belfast

Riccardo Scarpa

Durham University Business School

This research is funded by National Science Foundation grant #SES-0351946 and Queen's University, Belfast, under a cooperative research and development agreement with the United States Air Force Academy. The opinions expressed herein are solely those of the authors and do not necessarily reflect the views of the National Science Foundation or of the authors' respective institutions.

Abstract

In this paper we set up a three-stage experimental and theoretical framework to investigate strategic behaviour and design induced status quo bias in choice experiments. The research demonstrates that: (1) Repeated multiple choice experiments are not demand revealing. Thus, they do not generate reliable estimates of willingness to pay. (2) Most non-demand revealing choices are for the second-best option in a choice set. Consistent with the predictions of voting theory, these choices of the second-best option occur when there is an undesirable third option in a choice set. (3) As a result of the mathematics of combinatorial choice set design, the status quo option frequently occupies the second-best position in a choice set. (4) Experimental subject choices of the status quo in the second-best position are consistent with theoretical predictions derived from the mathematics of combinatorial choice set design. Although choice experiments as currently used in the field cannot be assumed to generate unbiased estimates of willingness to pay, this study demonstrates that the bias is of a predictable nature and direction.

Keywords: Choice experiments, public goods, non-incentive compatibility, status quo bias, choice set design, experimental economics

Introduction

In a typical choice experiment survey, respondents make a series of choices over sets of goods which vary in terms of the levels of their attributes and their costs. The results of these surveys enable researchers to estimate the potential economic benefits from the good or program being valued as well as consumers' willingness to pay (WTP) for various attributes of the good or program. Initially applied to problems in marketing (Louviere and Woodworth 1983), the methodology has been widely applied to estimate consumer preferences regarding health, transportation, and the environment (e.g. Ryan 2006, Hensher and Rose 2005, Adamowicz et al. 1998).

As the use of the method for policymaking has proliferated, so has the number of field and lab studies that examine whether the willingness to pay estimates generated by choice surveys are reliable in the sense that they accurately reflect consumers' underlying preferences. One of the most persistent findings of these studies is that of status quo bias. Status quo bias is defined as the status quo option being selected as the favorite alternative too frequently relative to what would be predicted if consumers' responses are consistent with their underlying preferences. Although many studies find that status quo bias is present, none of the experiments conducted to date is able to identify the exact source of the bias. This is because the potential sources of bias are confounded in the experimental designs. One group of studies (Carlsson and Martinsson 2001, Foster and Mourato 2002, Lusk and Schroeder 2004, Collins and Vossler 2009, Day and Pinto-Prades 2010, Taylor, Morrison, and Boyle 2010, and Vossler, Doyon, and Rondeau 2012) all use experimental designs that consist only of repeated choice elicitation formats. Although one might conclude from these studies that biased responses are more likely to occur in repeated choice settings, there is no one-shot control treatment to establish this conclusion. One study

(Lusk, Fields and Prevatt 2008) employs a strictly one-shot experimental design. Another study (List, Sinha and Taylor 2006) compares choices in one-shot and repeated settings, but the one-shot experiment is for a public good, while the repeated choice experiment is for a private good. Thus the effect of the choice setting on the reliability of preference estimates is confounded by the different incentive structures that are present in public and private goods environments. Consequently, although there is evidence that choice experiments may not generate reliable estimates of underlying willingness to pay, there is little evidence regarding *why* this is the case. The purpose of the research reported here is to fill this knowledge gap.

Insert Figure 1 about here.

This paper reports the results of an evolving inquiry into the reliability of willingness to pay estimates derived from choice experiments. The inquiry consists of three parts. First, an initial experiment reported in Part I whose results motivated the development of the theoretical model reported in Part II, followed by a second experiment (reported in Part III) to test the predictions of this theoretical model. Figure 1 summarizes the research design and identifies the main research question and results discovered at each stage of the inquiry.

The purpose of the experiment that began this inquiry to was to fill the gap in the literature identified above, by determining, what, if any aspect of a choice mechanism leads to non-demand revealing behavior. The experiment examines four choice experiment mechanisms: one-shot binary choice, repeated binary choice, one-shot multiple choice, and repeated multiple choice in an induced value setting to identify deviations from truthful preference revelation in each choice mechanism. The results demonstrate that of the four choice mechanisms tested, only the repeated multiple choice format generates unreliable estimates of underlying induced preferences. Both economic theory and the voting literature predict that this choice format will lead to strategic, non-demand revealing choices, in particular of the second-best option in a choice set. The data from the experiment are consistent with this prediction. Peculiarly, most of the second-best choices were for the status quo option, which was frequently in the second-best position in the choice set. Given that the choice sets were designed using standard methods of combinatorial choice set design, this result was surprising, and led to the conjecture that methods to create fractional factorial choice set designs result in many choice sets in which the status quo is the second-best option.

The theoretical model reported in Part II demonstrates that the frequent appearance of the status quo in the second-best position is not a result of methods to create fractional factorial choice set designs, but rather arises in the full factorial and persists when choice experiment designs based on fraction of the full factorial are created. The mathematical model demonstrates that the status quo option falls in the second-best position in a predictable fraction of choice sets. If subjects cast non-demand revealing votes for the second-best option, the observed behavior will look like status quo bias.

The second experiment, reported in Part III returns to the experimental laboratory to test these theoretical predictions using common methods to create fractional factorial designs for repeated multiple choice experiment surveys. The results are consistent with the theoretical predictions and demonstrate that repeated multiple choice experiments are not immune to the efficiency-bias trade-off that is present in other non-market valuation mechanisms.

The combination of the experimental and theoretical results demonstrates not only *that* biased responses to choice experiment surveys occur, but *why* they occur. Knowing the origins of the bias provides insight into how choice set designs, or methods for analyzing choice data, or both, can be modified to improve the usefulness of data derived from choice experiment surveys for policymaking. The remainder of this paper proceeds as follows. The next section summarizes the main result of the first experiment, which generates the hypotheses that are further developed in the theoretical model in Part II and tested in the experiment in Part III. Part IV discusses the overall findings and conclusions.

Part I. Experiment 1: How does choice mechanism design affect demand revelation in discrete choice experiments?

The purpose of Experiment 1 is to isolate the effects on demand revelation of two aspects of the choice mechanism – the number of options in the choice set and the number of choice occasions that subjects face – in a choice experiment. As noted above, previous experimental studies have failed to isolate the effects of these two design features on subject choices by failing to consider both one-shot and repeated versions of binary and multiple choice formats for the same type of good in the same experiment. We know from mechanism design theory that the incentives created by the structure of the choice mechanism affect the choices that subjects make, thus it is important to determine whether either of these foundational aspects (number of choice occasions and number of options in a choice set) of a choice experiment design affect the reliability of the resulting choice data.

An induced-value design (both here and in Part III) is necessary in order to cleanly identify differences between stated and actual preferences. Developed by Smith (1976), induced-value experiments create a set of "artificial" preferences for subjects by linking subjects' experimental earnings to the choices that they make in the experiment. Considered the gold standard in experimental economics, induced values allow us to observe what, if any, aspects of the choice experiment mechanism contribute to biased responses. A response is considered to be "biased" if the stated preference is not consistent with the underlying induced values of the subject. In the lab, if a subject chooses a response that does not maximize her experimental earnings, the choice is termed to be *not* demand-revealing, and hence, biased. In a field choice experiment survey, a non-demand revealing or biased choice would similarly be a choice of an option other than the respondent's most-preferred option though this is not observable as it is in the experimental laboratory.

We focus on the public goods environment, as this is the setting that is most relevant for environmental and social policy. The four methods for collecting choice experiment data studied in Experiment 1 are:

> One-Shot Binary Choice (OSB) – subjects face a single choice occasion in which they choose between two options, one of which is the status quo option. (8 groups of 9 subjects, 72 subjects in total)
>
> One-Shot Multiple Choice (OSM) – subjects face a single choice occasion in which they choose among three options, one of which is the status quo option. (8 groups of 9 subjects, 72 subjects in total)
>
> Repeated Binary Choice (RB) – subjects make a series of six choices. On each choice occasion, subjects choose between two options, one of which is the status quo option. (4 groups of 9 subjects, 36 subjects in total)
>
> Repeated Multiple Choice (RM) – subjects make a series of nine choices. On each choice occasion, subjects choose among three options, one of which is the status quo option. (4 groups of 9 subjects, 36 subjects in total)

The OSB treatment is the experimental control, as it is theoretically incentive compatible. This means that in this treatment, a subject can do no better than indicate her actual first choice as her stated preference. In the other treatments, which are versions of non-trivial voting mechanisms, it is a well-known result of the Gibbard-Satterthwaite theorem (Gibbard 1973, Satterthwaite 1975) and voting theory (Myerson and Weber 1993) that subjects may have an incentive to indicate a preference for an option other than his or her true first choice. In elections in which there are more than two candidates, voters sometimes have an incentive to vote for their second-best candidate in order to prevent other candidates that they dislike more from winning. This incentive depends on voters' beliefs about the distribution of preferences among the electorate. If voters believe that all candidates are equally likely to win, they are said to have uniform priors. Under uniform priors, a voter's best strategy is to cast vote for his or her most-preferred candidate. If a voter believes that the distribution of votes among candidates is not uniform, and in particular that a non-favored candidate is likely to win (a condition known as non-uniform priors), s/he may have an incentive to vote for a candidate other than the most-favored candidate.

Voting behavior in multi-candidate elections has been well-studied both in the laboratory and the field. In both settings, the evidence is consistent some voters strategically misrepresenting their preferences. The rate of strategic voting in these studies ranges from a low of 1.2% of voters to a high of over 50% (Forsythe et al. 1993, Fujiwara 2011, Kawai and Watanabe 2013). It is an empirical question whether subjects actually respond to the incentives present in these choice experiment mechanisms in a manner that is consistent with the theoretical and empirical literature on voting theory. The purpose of the experiment is to investigate which, if any of the choice experiment mechanisms can generate empirically demand-revealing results which would potentially generate reliable preference estimates for policy-making purposes.

Because many aspects of the design of Experiment 1 are similar to the design of Experiment 2, we discuss the design of Experiment 2 in Part III below, and relegate a detailed description of the Experiment 1 treatments to the supplemental materials, in order to focus the discussion below on the results of Experiment 1.

Insert Table 1 about here.

The main result of Experiment 1 is that the percentage of non-demand revealing choices is significantly greater in the Repeated Multiple Choice (RM) treatment than in either of the binary choice treatments (OSB and RB). This non-demand revealing behavior is consistent with a pattern of choices of the second-best option, which is frequently the status quo option. Table 1 reports the results of hypothesis tests showing that there is significantly more non-demand revealing behavior in the RM treatment than in the binary choice treatments.[1] As illustrated in Table S2 of the supplemental materials, subject fatigue (or other order effects such as learning, initial confusion or later stage fatigue) does not appear to be a useful explanation for the presence of these non-demand revealing choices, as the rate of non-demand revelation is consistent across the sequence of choices.

Insert Table 2 about here.

Table 2 reports the percentage of non-demand revealing (NDR) choices in the RM treatment by induced values, which shows that most non-demand revealing choices are for the second-best option in the choice set. Frequently, the second best option is the status quo option. 92% of the non-demand revealing choices are for the second-best option. In 7 of the 9 choice sets, the second-best option in the choice set is the status quo option. Given that the choice sets were created using standard methods of combinatorial choice set design, the result that 7 of the 9 choice sets have the status quo as the second-best option is surprising. Thus, it appears that it is the combination of repeated choices and choices among more than two options that generates non-demand revealing behavior in a choice experiment, that non-demand revealing behavior is driven by choices of the second-best option, and that the second-best option is frequently the status quo option. These results point to the following hypotheses which we examine in the remainder of this paper:

Hypothesis 1: In repeated multiple choice experiments, the mathematics of combinatorial choice set design result in a large number of choice sets in which the status quo is the second-best option.

Hypothesis 2: Consistent with the predictions of voting theory, non-demand revealing choices in a repeated multiple choice experiment will be choices of the second-best option, which is often the status quo option.

We examine hypothesis 1 by delving into the mathematics of combinatorial choice set design, and return to the experimental lab to examine hypothesis 2 in Experiment 2 in Section IV.

Part II. Theory: How do the mathematics of combinatorial choice set design affect the relative position of the status quo option in each choice set in a repeated multiple choice experiment?

The purpose of the theoretical inquiry reported in this section is to determine whether the frequent appearance of the status quo option in the second-best position in choice sets is an artifact of poor experimental design, or arises naturally out of the choice set creation process. We show that the latter is the case. In multiple choice sets derived from a full-factorial design, the second-best position in the choice set is frequently occupied by the status quo option. This is a result of the mathematics of combinatorial choice set design, and is also present when common methods to derive fractional factorial designs from the full-factorial design are employed.

We consider attribute balanced choice sets in which every option has the same number of attributes and every attribute has the same number of levels. The results below easily generalize to choice sets in which different attributes have different numbers of levels.[2] In choice sets in which each of m attributes have n levels, a full factorial design will contain n^m possible choice options which can be combined into choice sets. If $\pi \in [0,1]$ is the proportion of choice options that are preferred by an agent to the status quo[3], then πn^m choice options will be preferred to the status quo, and $(1-\pi)n^m$ choice options will not be preferred to the status quo. In a multiple choice set containing two options and the status quo, there are $n^m(n^m-1)$ possible choice sets that consist of two distinct options and the status quo. Of these choice sets, $(\pi n^m)(\pi n^m - 1)$ of them will have the status quo as the worst option in the choice set, and $[(1-\pi)n^m][(1-\pi)n^m - 1]$ of them will have the status quo as the best option in the choice set. This leaves

$$n^m(n^m - 1) - \{\pi n^m(\pi n^m - 1) + [(1-\pi)n^m][(1-\pi)n^m - 1]\}$$

choice sets that contain the status quo as the second-best option. It is possible to show that in the limit (as either $n \to \infty$ or $m \to \infty$), the ratio of this number to the total number of choice sets is $2\pi(1-\pi)$. It is also possible to show that in the limit, the fraction of choice sets in the full factorial in which the status quo option is the first-best or second-best option will be $(1-\pi^2)$. Figures 2 and 3 illustrate these results. The complete derivation of these results is provided in the supplemental materials.

Insert Figures 2 and 3 about here.

The results in Figure 3 indicate that the status quo is likely to land in the first-best or second-best position in a large fraction of choice sets. In a choice experiment in which half the options are preferred to the status quo and half are not ($\pi = 0.5$), in the limiting case, the status quo will reside in the first-best or second-best slot in 75% of the choice sets constructed from the full factorial. For smaller values of n or m and $\pi = 0.5$, the fraction of choice sets in which the status quo is first- or second-best exceeds 75%, and approaches the limiting value of 75% relatively quickly. For example, as illustrated in Table 3, in a choice experiment with four attributes each with four levels, 75.1% of the choice sets in the full factorial will contain the status quo in the first- or second-best position.

Insert Table 3 about here.

Most field choice experiment surveys do not employ a full factorial design, as such a design would require respondents to answer a very large number of choice questions. We employ simulation methods to investigate to what extent the mathematical predictions for the full factorial design are relevant to frequently used fractional factorial designs. We employ two techniques to obtain fractional factorial designs: orthogonality on the attribute level differences (Kanninen 2002; Street et al. 2005) and utility balance (Huber and Zwerina 1996). Using Ngene v. 1.2., a specialized software for choice set design, we derive designs for experiments involving 12 choice tasks using choice sets with three attributes (A, B, and Cost) with two levels each (1 or 2 units of A, 1 or 3 units of B, and Costs of 5 or 10). The status quo option contains one unit of attribute A and one unit of attribute B at no cost. In implementing the search for optimal designs under both criteria, we strove to preserve attribute level balance and the presence of tradeoffs in all choice tasks.

Insert Figures 4 and 5 about here.

After generating the choice experiment designs, we vary the marginal values of the attributes in order to vary the value of π. Table S3 reports the values of π employed in the simulation and the marginal values of the attributes associated with each value of π. It is then possible to calculate the utility from each option in a choice set using the equation

$$\text{Utility} = (\text{Value of A} \times \text{Units of A}) + (\text{Value of B} \times \text{Units of B}) - \text{Cost}$$

and to determine the relative position of the status quo in each choice set. Figures 4 and 5 report the results of the simulation for 12 choice tasks. The results of these simulations indicate that the mathematical predictions based on the full factorial are useful as approximations for reductions from the full-factorial design.[4] It remains to be seen whether this result is robust to more complex choice experiment designs that contain more attributes and/or more attribute levels.

Insert Table 4 about here.

Given that the status quo will often appear in the second-best position in choice sets in a repeated multiple choice experiment, what incentives do subjects have to choose this second-best option? In a repeated multiple choice experiment, if subjects believe that each option is equally likely to be chosen (uniform priors), then their optimal choice is to select their most-preferred option. If, however, some subjects believe that an option that they really dislike is likely to capture a large number of votes from other subjects (non-uniform priors), then the incentive exists for them to vote for their second-best option if they think it has a better chance of preventing the third option from winning. Do these incentives to vote strategically exist in a repeated multiple choice experiment? The answer is yes. To see why, it is useful to consider a numerical example. We employ the orthogonal on the differences (OOD) choice set with 12 choice tasks from the simulations above to examine the fraction of respondents for whom each option is their most-preferred option. Table S4 reports the OOD solution for this choice set design, for which D-optimality is 100%. Each of these choices is combined with the status quo option to create twelve multiple choice sets. To calculate the fraction of respondents for whom each option is the preferred choice, we assume that subjects in the sample have values of π that are uniformly distributed between 0 and 0.75 as illustrated in Table S3. Table 4 reports results of this exercise. If we consider this table to represent respondents' priors about other respondents' preferences, then non-uniform priors are the norm. It is in exactly these types of cases that incentives to cast votes for second-ranked options exist. Given the mathematics of combinatorial choice set design, the second-ranked option will frequently be the status quo option.

How would this work in the field? As illustrated in Figures 4 and 5, the two fractional factorial methods (OOD and UBAL) present the status quo as the second best option in many choice sets when π takes low to medium values (0.33 to 0.5 in Figure 4) and as either the first or second best option on all choice occasions for values of π below 0.4 (in Figure 5). The fact that the status quo never appears as the worst option in OOD and UBAL multiple choice designs for a substantial proportion of the sample with low values of π is strongly indicative to such low π subjects that it is a safe choice likely to obtain mores choices than the worst option in their choice set. In real field situations low values of π are likely to arise for a substantial proportion of the sample if they regard the price vector as high relative to the utility of non-status quo choices. Thus, the mathematics of combinatorial choice set design combined with the incentives in a repeated multiple choice mechanism for a public good can result in non-demand revealing choices of the second-best option, which will often happen to be the status quo option. We now return to the lab to examine whether behavior is consistent with the predictions of this theoretical model.

Part III. Do subjects in a repeated multiple choice experiment behave in a manner consistent with the predictions of voting theory combined with the theoretical model from Part II?

A. Experiment 2 Design

Experiment 2 is designed to create a laboratory simulation of a field multiple choice experiment for a public good in order to test the predictions of the mathematical model from Part II combined with the predictions of voting theory. In the experiment, each subject faces a series of 12 choice sets. In each choice set, the subject must choose from three options – two alternatives and the status quo. In order to properly incentivize the experiment as required by induced value theory, subjects' choices must be made binding. After subjects make all 12 choices, the binding ballot is determined by rolling a 12-sided die. The binding ballot is then counted to determine the winning option. Two-way ties are resolved via a coin toss. Three-way ties are resolved via the toss of a 6-sided die. Each subject's ballot is returned to him or her, and they compute their earnings in dollars for the winning option using the equation:

$$\text{Earnings} = 6 + (\text{Value of Attribute A} \times \text{Units of Attribute A}) + (\text{Value of Attribute B} \times \text{Units of Attribute B}) - \text{Cost}$$

Subjects receive their experimental earnings in cash prior to departing the experimental session.

As indicated above, each choice option has two attributes and a cost. Each attribute and the cost have two possible levels. Thus, the choice set design employs a 2^3 factorial design. In order to test whether some choice set creation methodologies result in choice experiments which are more prone to non-demand revelation, we employ three methods to create fractional factorial choice set designs: orthogonality on the attribute level differences (OOD) (Kanninen 2002; Street, Burgess, and Louviere 2005), utility balance (UBAL) (Huber and Zwerina 1996), and randomly generated choice sets (RAND).

In an OOD choice set in which each attribute has only two levels, the smallest possible fractional factorial choice experiment resulting from a 2^3 full-factorial design contains twelve choice sets containing two options and the status quo. In order to maintain consistency across treatments, the UBAL and RAND designs also contain twelve choice sets. In a field choice experiment using utility-balanced choice sets, the experimental designer would have to make some assumptions regarding which choice options are likely to be considered better or worse by respondents in order to create utility balanced choice sets. In an induced-value setting, subjects' utility for each choice option is known via the induced value function. Thus, we create a different set of utility-balanced choice sets for each set of induced values. The supplemental materials report the choice set designs for the OOD, UBAL, and RAND methods in Tables S4 through S8. The order of choice sets is randomized across subjects, so that on any given choice occasion, different subjects may be choosing from a different set of choice options.

Subjects know all of the rules of the game at the beginning of the experiment. Each subject knows that other members of the group may face a different set of choice options on any given ballot than s/he does, and knows that different subjects may also have different marginal values for the attributes than s/he does. No subject knows the overall distributions of marginal values or possible choice combinations.

Insert Table 5 about here.

In order to test the mathematical predictions of Part II, the experimental design incorporates three sets of induced values, corresponding to low, medium and high values of π of 0.125, 0.50, and 0.75. When $\pi = 0.125$, 12.5% of the choice options are preferred to the status quo option. Similarly, when $\pi = 0.50$ and 0.75, 50% and 75% (respectively) of the choice options are preferred to the status quo. These values of π vary individually by subject. In the experiment, the variation in π is generated by assigning subjects different marginal values of the attributes of the choice options, as would also be the case in a field choice experiment. Table S2 reports the marginal values of attributes A and B that correspond to the values of π used in the experiment.

Table 5 reports the predicted fraction of choice sets in which the status quo is second-best or first- or second-best in the full factorial design using both the non-limiting case for $n = 2$ and $m = 3$ and the limiting case derived in Part II, as well as the actual fraction of choice sets in each fractional factorial design in which the status quo is second-best or first- or second-best for each combination of π and choice set design. The fraction of the time that the status quo appears as the second-best or first- or second-best option in each choice set design is consistent with the predictions of the mathematical model.

The combination of three methods to create fractional factorial choice experiments and three sets of induced values results in nine possible experimental treatments. Within an experimental session, which could contain up to 27 subjects, subjects were randomly assigned to a group of nine who participate in one version (OOD, UBAL or RAND) of the choice experiment.[5] Within a group of nine, subjects were randomly assigned their induced values. Subjects were blinded to the method by which their choice sets were created, the induced values of the other individuals in their group, and the overall distribution of induced values. The moderator of a given group of nine was also blinded to the distribution of induced values within that group. Two hundred and seventy student volunteers from a U.S. university participated in the experiment, resulting in a sample of 30 subjects for each choice set design-induced value combination. Given that each subject makes 12 choices, this results in a total of 360 observations per choice set design-induced value combination, or 1080 observations for each of the choice set design methodologies. The experiment took 40-45 minutes to complete, and average experimental earnings were in the range of $15-$20.[6] After completing the experiment, subjects filled out a brief demographic questionnaire. In order to preserve subject anonymity, the demographic survey was limited to collecting information about gender, class year, and major field of study. Major field of study was aggregated to the level of Technical Majors (Engineering and Basic Sciences majors), Non-Technical majors (Social Sciences and Humanities majors), and Undeclared, to preserve anonymity.[7] Subjects were free to leave any or all demographic questions blank. Table S9 summarizes the demographic characteristics of the experimental subjects. The sample is representative of the overall student body in terms of gender and major field of study. Sophomores are heavily over-represented in the sample, most likely because a great deal of subject recruiting happened in the core economics course that is a required course for all students. Most students take this class during their sophomore year. In general, there were no differences in subject behavior by demographics, so we do not discuss them here. Table S10 reports the rates of demand revelation by subject demographic.

B. Experiment 2 Results

This experiment is designed to test the following hypotheses:

H1: Most non-demand revealing choices are for the second-best option in the choice set. Very few non-demand revealing votes are for the worst option in the choice set.

H2: Consistent with a model of strategic voting, subjects make non-demand revealing choices for the second-best option when there is a very bad third option in the choice set.

H3: The fractions of non-demand revealing choices of the status quo as the second-best option are consistent with the mathematical predictions based on the value of π reported in Table 5.

H4: There is no difference in the rate of demand revealing behavior across methods (OOD, UBAL, or RAND) to reduce the number of choice sets from the full factorial design.

H5: There is no difference in the rate of demand revealing behavior across subjects' values of π (the proportion of overall choices preferred to the status quo).

For all hypothesis tests reported in the next section, standard errors are clustered at the individual subject level to account for the fact that each subject makes 12 choices.

Insert Figure 6 about here.

Result 1: Most non-demand revealing votes are for the second-best option. Consistent with models of strategic voting, these votes tend to occur when there is a bad third option in the choice set. Very few non-demand revealing votes are for the worst option in the choice set. The evidence confirms hypotheses 1 and 2.

One hundred twenty-two subjects (45% of all subjects) cast at least one non-demand revealing vote during the experiment. Figure 6 reports the frequency of total number of non-demand revealing votes cast by individual subjects. The vast majority of subjects who cast non-demand revealing votes cast one or two non-demand revealing votes out of a possible 12. Conditional on casting a non-demand revealing vote, the mean number of non-demand revealing votes is 1.22, with a standard deviation of 2.01.

Insert Table 6 about here.

Table 6 reports the fraction of non-demand revealing votes for the second-best and worst options, by type of choice set design and induced value, with standard errors clustered at the subject level. These results reveal rates of strategic voting that are lower than observed in previous lab studies (e.g. Felsenthal, Rapoport and Maoz 1988; Forsythe et al. 1993), but higher than observed in the field (Kawai and Watanabe 2013). Previous lab studies of strategic voting have provided subjects with complete information about the distribution of other voters' preferences, giving full information about non-uniform priors and making the incentives for strategic voting completely transparent. In this experiment, subjects have complete information about their own preferences as in other lab studies, but must form priors about the distribution of preferences of other members of the group, as in field choice experiment studies. Thus, given the hybrid nature of the experimental design, it is heartening to observe that rates of voting for the second best alternative are consistent with and midway between previous results from the lab and the field.

In each treatment, over 70% of non-demand revealing votes are for the second-best option. Conditional on a vote being non-demand revealing, there are no significant differences in the fractions of non-demand revealing votes for the second-best option across choice set design treatments ($0.55 \leq p \leq 0.90$). Subjects with low values of π are significantly more likely to cast a non-demand revealing vote for the second-best option, and significantly less likely to cast a non-demand revealing vote for the worst option than subjects with a value of π equal to 0.75 ($p = 0.06$). There are no other significant differences in the nature of non-demand revealing behavior across subject types. These results are consistent with 1, and demonstrate that non-demand revealing behavior is not random, but rather follows a pattern consistent with models of strategic voting.

Insert Figure 7 about here.

Including the status quo option, there are only nine possible choice options in the choice sets used in Experiment 2. . Figure 7 reports the ranking out of 9 of the worst pay-off option in the choice set on those ballots for which non-demand revealing votes for the 2^{nd} best option were cast. When subjects cast a non-demand revealing vote for the second-best option, 65% of the time the ballot contained a third option which ranked 7^{th}, 8^{th}, or 9^{th} out of the possible nine choice options. Seventy-nine percent of these ballots contained a third choice option which ranked in the bottom half of the ranking of choice options. Thus it appears that non-demand revealing votes for the second-best option occur when there is a bad third option on a ballot that a subject is trying to avoid, consistent with models of strategic voting and hypothesis 2.

Insert Table 7 about here.

Result 2: Fractions of non-demand revealing choices of the status quo as the second-best option in the choice set are generally consistent with the mathematical predictions based on the value of π from Part II. The evidence is consistent with hypothesis 3.

Table 7 repeats the first five rows of data from Table 5, and adds the fraction of second-best votes for the status quo option observed in the experiment for each choice set design. In the experiment, the proportion of second-best votes for the status quo should reflect the number of times the status quo appears as the second-best option in a choice set. These proportions arise from the mathematical predictions that are derived in Part II. Consider the case of $\pi = 0.75$ as an example. In the OOD choice sets for subjects with $\pi = 0.75$, half the choice sets have the status quo in the second-best position and half do not. Thus, we should expect to see half of the second-best votes cast by $\pi = 0.75$ subjects in the OOD experiment to be for the status quo option. In actuality, 40% of the second- best votes cast by $\pi = 0.75$ subjects in the OOD experiment were for the status quo option. It is possible for subjects to cash a second-best vote for the status quo as second-best option more often than predicted by the mathematical model from Part II, or less often, depending on which choice occasions the subjects choose to cast second-best votes. If they are casting second-best votes more often on choice occasions in which the status quo is not in the second-best position, the observed fraction of second-best votes will be less than predicted by the mathematical model. If they are casting second-best votes more often on choice occasions where the status quo is in the second-best position, then the observed fraction of second-best votes for the status quo option will be greater than predicted by the mathematical model. Both outcomes are observable in Table 7. Which choice occassions subjects choose to cast second-best votes on depends on a subject's priors and how bad the third option is for them, which varies from choice set to choice set and by value of π. The exception to this are the OOD choice sets for $\pi = 0.50$ subjects. Since all of the choice sets contain the status quo in the second-best position, all second-best votes will necessarily be for the second bt option. The key message of Table 7 is, with the exception of the UBAL treatment, the fraction of second-best votes for the status quo option is generally consistent with the mathematical predictions from Part II. In general, there is an inverted U-shape between the value of π and the fraction of second-best votes for the status quo option. This table supports our conclusion that the fraction of non-demand revealing votes that are for the status quo option are of a predictable nature and direction..

In the OOD treatment, the fraction of subjects casting non-demand revealing votes for the status quo option as the second-best option is significantly greater for subjects with $\pi = 0.50$ than for subjects with $\pi = 0.125$ ($p < 0.0001$) or $\pi = 0.75$ ($p = 0.0004$), a result that is also consistent with the mathematical model of Part II, since the fraction of choice sets in which the status quo is second-best is maximized when $\pi = 0.5$. High π ($\pi = 0.75$) subjects also cast significantly more non-demand revealing votes for the status quo as second-best than low π ($\pi = 0.125$) subjects ($p = 0.02$). The same pattern is present for the RAND treatment. Subjects with $\pi = 0.50$ cast the most non-demand revealing votes for the second-best in favor of the status quo ($p < 0.0001$ for $\pi = 0.125$ vs. $\pi = 0.50$, $p = 0.04$ for $\pi = 0.50$ vs. $\pi = 0.75$), and subjects with high π ($= 0.75$) cast more non-demand revealing votes in favor of the status quo as second best option than subjects with low π ($= 0.125$) ($p = 0.0007$). The UBAL treatment does not follow the pattern as closely. Although subjects with high π are significantly more likely to cast non-demand revealing votes in favor of the status quo as the second best option ($p = 0.05$ for $\pi = 0.125$ vs. $\pi = 0.75$, $p = 0.04$ for $\pi = 0.50$ vs. $\pi = 0.75$), there are no significant differences in the fraction of non-demand revealing second-best votes for the status quo for the two lower values of π ($p = 0.53$). With the exception of the UBAL treatment, the results are consistent with hypothesis 3. Most importantly, for OOD choice sets, which are the most commonly used choice set design in field settings, the results are strongly consistent with the mathematical model presented in Part II.

Insert Table 8 about here.

Result 3: There are differences in the rate of demand revelation by type of choice set design, value of π. Thus, hypotheses 4 and 5 are rejected.

Table 8 reports the fraction of votes that are demand revealing by choice set design type and value of π. The rate of demand revelation is significantly less in the UBAL treatment than in the OOD treatment ($p = 0.02$). There are no significant differences in the rates of demand revelation between the OOD and RAND treatment, or between the UBAL treatment and the RAND treatment. Thus, there is weak evidence that the type of method used to create fractional factorial choice set designs from the full factorial affects the rate of demand revelation one obtains in a field choice experiment survey, with utility-balanced designs generating less demand-revealing results. This result implies that hypothesis 4 should be rejected.

Subjects who prefer very few choice options to the status quo ($\pi = 0.125$) are significantly less likely to cast demand revealing votes than subjects who prefer a higher fraction of options to the status quo ($\pi = 0.50$ or 0.75, $p = 0.03$ for $\pi = 0.125$ vs. 0.50 and $p = 0.02$ for $\pi = 0.125$ vs. $\pi = 0.75$). Thus, one's propensity to cast a non-demand revealing vote seems to be significantly related to how one values the other choice options relative to the status quo, evidence that implies that hypothesis 5 should be rejected. Survey respondents with a low π prefer the status quo to most of the other choice options available (all but one, in the case of this experiment), and have many bad options in the choice sets that they face. Thus, these subjects may be more likely to have non-uniform priors about the distribution of the other subjects' votes, and be more likely to cast non-demand revealing votes in an effort to avoid these bad options being imposed on them by others for whom they may be preferred choices.

Day and Pinto-Prades (2010) demonstrate a similar result by deliberate framing, which either used an option with the same attributes appearing in a choice set at a higher price, or an option with the same price but worse attributes appearing in the choice set to create "a bad third choice." Both of these characteristics would reduce the desirability of this choice option, and would result in it having a lower ranking among the choice options in the full or fractional factorial design. As a choice option's ranking relative to the status quo falls, the value of π falls. Thus, creating these framing effects is equivalent to lowering the value of π for a given choice set design. Lowering the value of π for a subject or group of subjects by introducing framing effects will result in more bad options in a choice set, and less demand revelation.

3. Alternative Explanations for Observed Behavior

There are three explanations of subject behavior which are observationally equivalent with the data from Experiment 2. The first is strategic voting. The second is errors in optimization, as suggested by Collins and Vossler (2009). Errors in optimization are a plausible explanation for the pattern in our data, if these errors are more likely to occur on choice occasions on which subjects eliminate the bad third option quickly, and then struggle to choose between the top two options.

Such optimization errors could result from neurologic choice processes, which is the third explanation for the pattern observed in our data. Krabich and Rangel (2011) find that fMRI data of subjects making binary choices between alternative snack foods is consistent with an optimizing model of choice with error. Such choice errors are more likely to occur when the options are similar in terms of attributes. Thus, neurologic choice processes may lead to choices of the second-best option consistent with our data if subjects rule out the bad third option quickly and focus on the more difficult choice between the top two options. Although observationally equivalent, only a strategic voting model would predict more non-demand revealing choices by low-π subjects. Therefore, of the three explanations, the experimental data is most consistent with a model of strategic voting. Whether as a result of strategic behavior, errors in optimization, neurologic choice processes, or some combination of the three, the experimental results demonstrate that the pattern of non-demand revealing choices will be consistent with the predictions of a model based on the mathematics of combinatorial choice, which predicts the fraction of second-best choices can be preponderantly for the status quo option.

Part IV. Discussion and Conclusion

This paper demonstrates that the incentives of a multiple choice mechanism for a public good interact with the mathematics of combinatorial choice set design to generate the potential for biased responses in a repeated multiple choice experiment that are of a particular type and direction. Non-demand revealing responses to the choice experiment survey are likely to (a) be choices of the second best option; (b) be more common for subjects who prefer the status quo to most of the choice options (either because they do not value the attributes highly, or they do not think that the improved attributes are worth the cost); (c) occur when there is a bad third option in the choice set; and (d) be a choice of the status quo option a predictable percentage of the time. These effects are robust across choice set design methodologies, and demonstrate that the problem is not one of modifying the choice set design.

When combined with recent work by Vossler, Doyon, and Rondeau (2012) and Petrolia and Interis (2013) these results might point to the conclusion that researchers should reconsider the usefulness of repeated trinary choice surveys in estimating preferences for public environmental goods and services. Although Vossler, Doyon, and Rondeau (2012) demonstrate that a repeated binary choice mechanism *could* be demand-revealing if respondents treat each choice question independently, Carson, Chilton, and Hutchinson (2009) show that a repeated binary choice setting *is* demand-revealing if respondents treat each choice question independently. They also show that the mechanism is not demand-revealing if the choice questions are not treated as independent. Petrolia and Interis (2013) argue that in a field setting, it may not be feasible to create a structure that induces respondents to treat choice questions independently, a point which Carson, Chilton, and Hutchinson (2009) and Day and Pinto-Prades (2010) make as well. Solutions to the problems of the mechanism must then either focus on how to best use the flawed information that the mechanism generates, or on finding new mechanisms with improved incentive properties.

Information from repeated multiple choice mechanisms may still be useful if researchers can develop ways to identify when respondents are choosing the second best option in the choice set. Choices of the second best option still contain information that is useful in estimating willingness to pay. It may be possible to identify from field data which choices are likely to be of the second-best option. In addition, researchers might consider developing econometric methods to estimate π as a separate parameter in a mixed logit model. This may enable them to determine which subjects are more likely to be choosing the second-best option in a choice set, and on which choice occasions these choices are more likely to occur.

Efficiency gains from repeated multiple choice mechanisms come from two sources. The first is from the additional information gained per respondent by having respondents make repeated choices. The second is from the addition of an option in the choice set. Adding an option to the choice set results in an additional utility difference in the likelihood function, which (assuming that responses are demand-revealing) results in more accurate preference estimates. It may be possible to modify multiple choice experiment surveys to maintain the efficiency benefits that they generate, while improving the incentive properties to avoid systematic biases in stated preferences. Mechanisms such as worst choice, or best-worst choice (Scarpa et al. 2011) hold some promise. Such modifications should be rigorously tested in a laboratory setting before field implementation to ensure that they are capable of generating robust and reliable estimates of willingness to pay that are not subject to strategic behavior.

References

Adamowicz W, Boxall P, Williams M, Louviere J (1998), Stated preference methods for measuring passive use values: choice experiments and contingent valuation, Am J Agric Econ 80(1): 64-75

Carlsson F, Martinsson P (2001), Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment, J Environ Econ Management 41(2): 179-192

Carson KS, Chilton SM, Hutchinson, WG (2009) Necessary conditions for demand revelation in double referenda, J Environ Econ Manage 57(2):219-225

Collins JP, Vossler CA (2009) Incentive compatibility tests of choice experiment value elicitation questions, J Environ Econ Manage 58(2):226-235

Day B, Pinto Prades J-L (2010), Ordering anomalies in choice experiments, J Environ Econ Manage 59(3):271-285

Felsenthal DS, Rapoport A, Maoz Z (1988), Tacit cooperation in three alternative non-cooperative voting games: A new model of sophisticated behavior under plurality procedure, Elect Stud 7: 143-161

Forsythe R, Myerson RB, Rietz TA, Weber RJ (1993), An experiment on coordination in multi-candidate elections: The importance of polls and election histories, Social Choice and Welfare 10(3): 223-247

Foster V, Mourato S (2002), Testing for consistency in contingent ranking experiments, J Environ Econ Manage 44(2): 309-328

Fujiwara T (2011), A regression discontinuity test of strategic voting and Duverger's Law, Q J of Polit Sci 6: 197-233.

Hensher DA, Rose J (2005), Respondent behavior in discrete choice modeling with a focus on the valuation of travel time saving, J Transportation and Statistics 8(2): 17

Huber J, Zwerina K (1996), The importance of utility balance in efficient choice designs, J Marketing Res 33(3):307-317

Gibbard A (1973), Manipulation of voting schemes: a general result, Econometrica: 587-601

Kanninen BJ (2002) Optimal design for multinomial choice experiments, J Marketing Res 39:214-217

Krajbich, I, Rangel A (2011), Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions, Proc Nat Acad Sci 108(33): 13852-13857

Kawai K, Watanabe Y (2013), Inferring strategic voting, Am Econ Rev 103(2): 624-662

List JA, Sinha P, Taylor MH (2006), Using choice experiments to value non-market goods and services: Evidence from field experiments, Adv Econ Analysis Pol 6(2):1-37

Louviere JJ, Woodworth G (1983), Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data, J Marketing Res 20: 350-367.

Lusk JL, Fields D, Prevatt W (2008), An incentive compatible conjoint ranking mechanism, Am J Agric Econ 90(2): 487-498

Lusk JL, Schroeder TC (2004), Are choice experiments incentive compatible? A test with quality differentiated beef steaks, Am J Agric Econ 86(2): 467-482

Myerson RB, Weber RJ (1993), A theory of voting equilibria, Am Pol Sci Rev 87: 102-114

Petrolia D, Interis M (2013) Should we be using repeated choice surveys to value public goods? Association of Environmental and Resource Economists Newsletter, November: 19-25

Ryan M, Netten A, D Skatun, P Smith (2006), Using discrete choice experiments to estimate a preference-based measure of outcome – an application to social care for older people, J Health Econ 25: 927-944

Satterthwaite MA (1975), Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions, J Econ Theory 10(2): 187-217

Scarpa R, Notaro S, Louviere JJ, Raffaelli R (2011), Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons, Am J Agric Econ 93(3):813–828

Smith V (1976), Experimental economics: induced value theory, Am Econ Rev: 274-279

Street DJ, Burgess L, Louviere JJ (2005) Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments, Int J Res Marketing 22:459-470

Taylor LO, Morrison MD, Boyle KJ (2010), Exchange rules and incentive compatibility of choice experiments, Environ Resource Econ 47(2):197-220

Vossler CA, Doyon M, Rondeau D (2012), Truth in consequentiality: Theory and field evidence on discrete choice experiments, Am Econ J: Microecon 4(4):145-171

Notes

[1] Standard errors are calculated assuming that the data are clustered at the level of an experimental session. An experimental session is a group of nine subjects. The method of preserving anonymity in the experimental design did not permit clustering of standard errors at the individual subject level. Note that assuming this form of clustering is a conservative approach, as it results in fewer clusters (and larger standard errors) than would result if standard errors were corrected for clustering at the individual subject level. Thus, any results that are found to be significant using this method to correct for clustering are sure to be significant when using a method which allows for more clusters.

[2] A proof of this result is available from the authors upon request.

[3] We assume a single value of π for all agents. The result is robust to relaxing this assumption in favor of the more realistic assumption that different individuals have different values of π, and if some average value of π for the respondent pool is assumed.

[4]A simulation with 20 choice tasks (available from the authors) generates nearly identical results.

[5] If a number of subjects that was not divisible by nine showed up to a session, the unassigned subjects were invited to sign up to participate in different session at a later time.

[6]Weekly student take-home pay ranges from $65-$100. Thus, an additional $15-$20 for 45 minutes to one hour of time represents a significant addition to a student's weekly income. Given this, there is no doubt that the experimental earnings were salient to the subjects.

[7] Copies of all experimental instructions and the post-experimental demographic survey are available from the authors upon request.

Experiment 1
How does choice mechanism design affect demand revelation in discrete choice experiments?

Key Design Variables:
One-Shot vs. Repeated Choice
Binary vs. Multiple Choice

Main Results:
1. Repeated multiple choice Experiments are not demand revealing.
2. Most non-demand revealing choices are for the second-best option in the choice set.
3. Frequently, the second-best option in the choice set is the status quo option.

Theoretical Model
How do the mathematics of combinatorial choice set design affect the relative position of the status quo option in each choice set in a repeated multiple choice experiment?

Main Results:
1. There is a predictable, mathematical relationship between the relative ranking of the status quo in choice options and the position of the status quo in three-option choice sets generated from a full-factorial design.
2. The predictions of the mathematical model are valid for fractional factorial choice set designs that are reductions from the full factorial design.

Experiment 2
Do subjects in a repeated multiple choice experiment behave in a manner consistent with the predictions of voting theory combined with the theoretical model?

Key Design Variables:
Method of Creating Fraction Factorial Choice Set Designs
Relative Ranking of the Status Quo among Choice Options

Main Results:
1. Repeated multiple choice experiments are not demand revealing.
2. Consistent with voting theory, most non-demand revealing choices are for the second-best option in a choice set.
3. Consistent with voting theory, these non-demand revealing choices for the second-best option occur when there is a very bad third option in a choice set.
4. The fraction of second-best choices for the status quo option is consistent with the predictions of the theoretical model.

Figure 1. Evolution of the Research Agenda

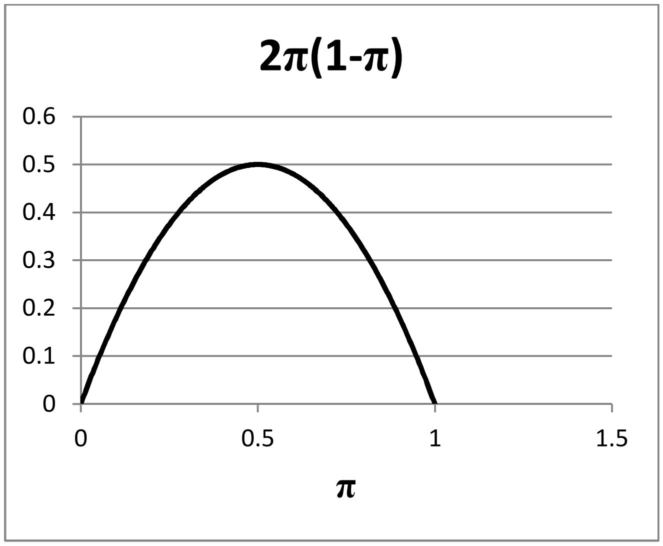

Figure 2. Limiting Proportion of Choice Sets in which Status Quo is Second-Best as a Function of the Proportion of Choice Options Preferred to Status Quo in the Full Factorial Design (π)

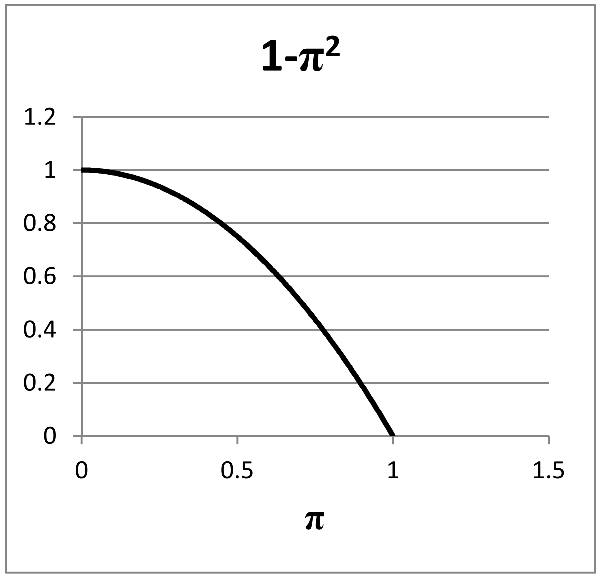

Figure 3. Limiting Proportion of Choice Sets in which the Status Quo is First or Second-Best as a Function of the Proportion of Choice Options Preferred to the Status Quo in the Full Factorial Design (π)

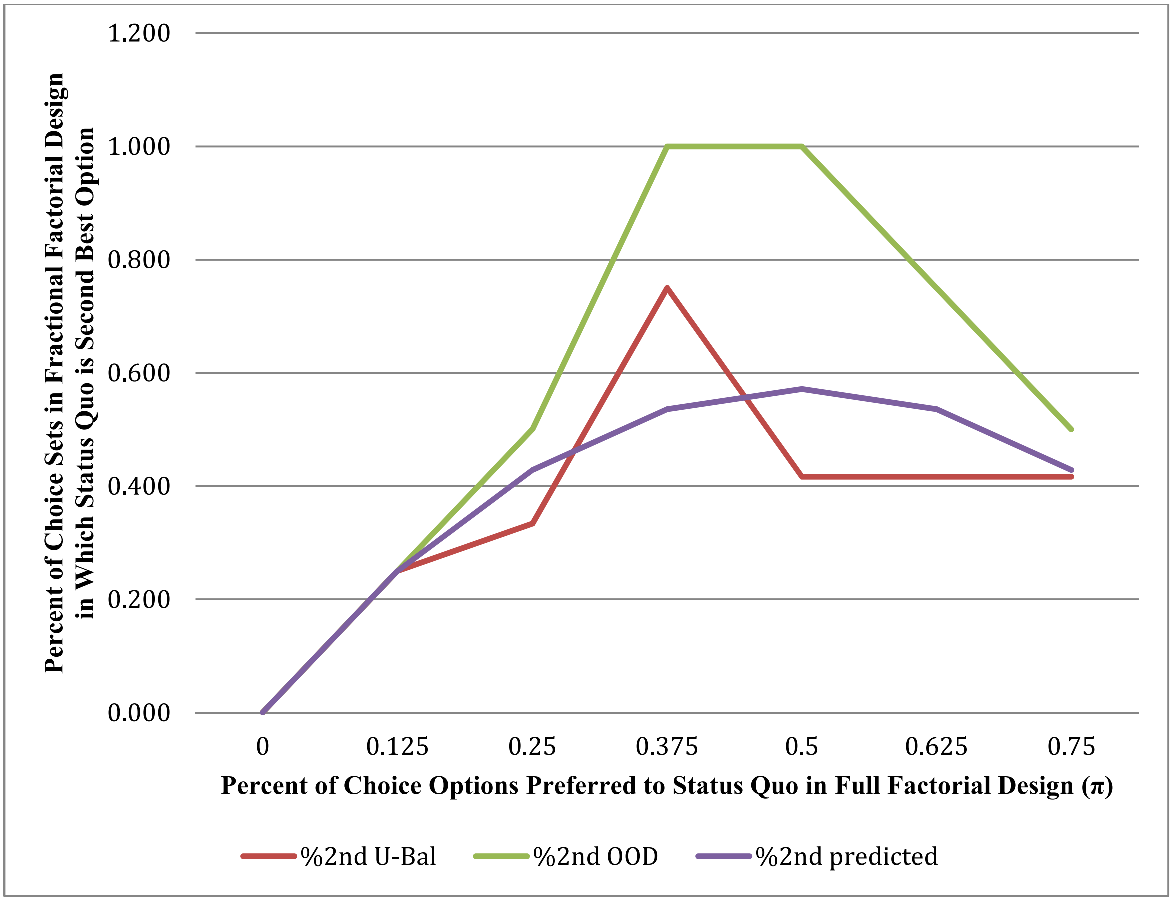

Figure 4. Predicted and Actual Proportion of Choice Sets in Which the Status Quo is Second-Best as a Function of the Fraction of Choice Options Preferred to the Status Quo (π) for Orthogonal on the Differences (OOD) and Utility Balanced (UBAL) Choice Experiments with 12 Choice Tasks

(prediction is based on the full-factorial design)

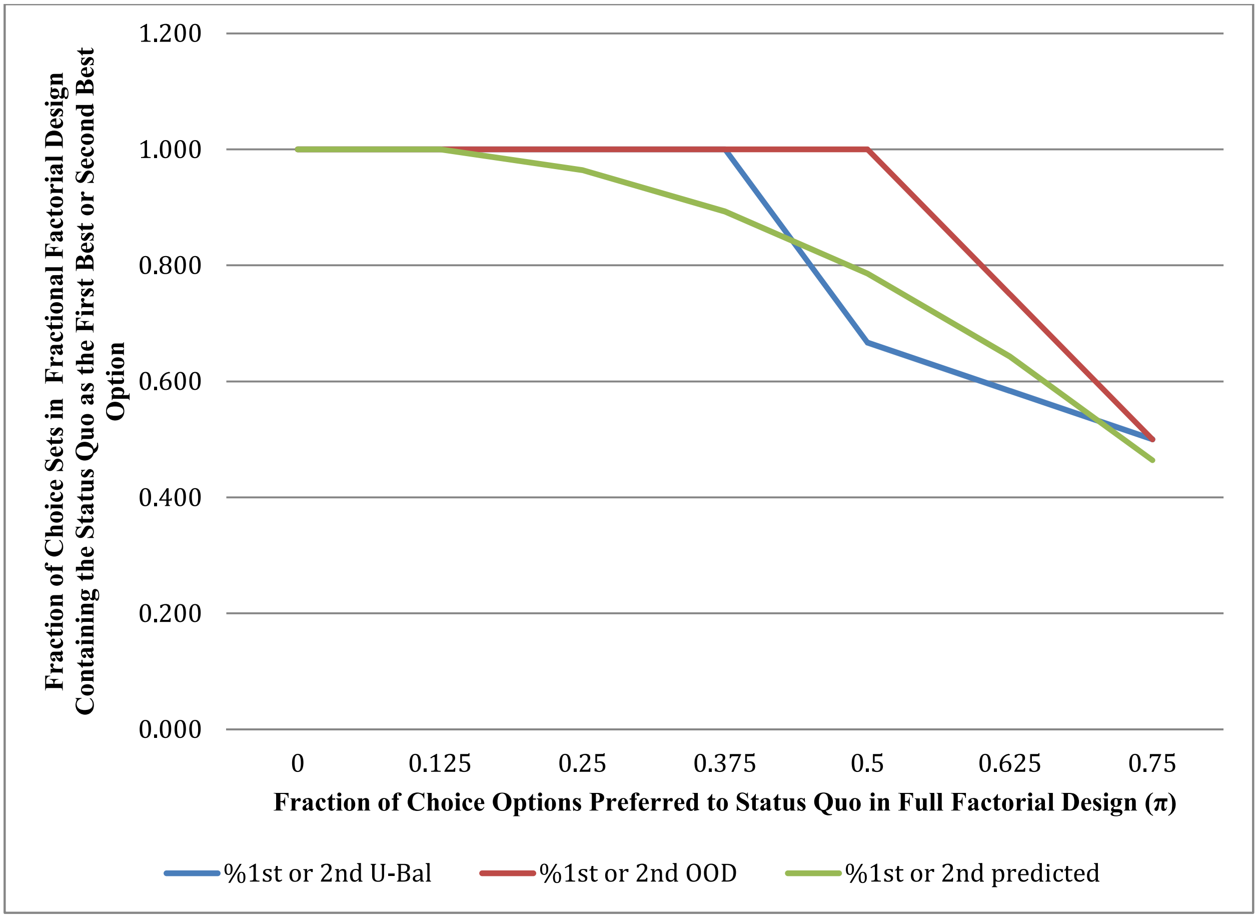

Figure 5. Predicted and Actual Proportion of Choice Sets in Which Status Quo is First-Best or Second-Best Option in the Choice Set as a Function of the Fraction of Choice Options Preferred to the Status Quo in the Full Factorial Design (π) for Orthogonal on the Differences (OOD) and Utility Balanced (UBAL) Choice Experiments with 12 Choice Tasks

(prediction is based on the full-factorial design)

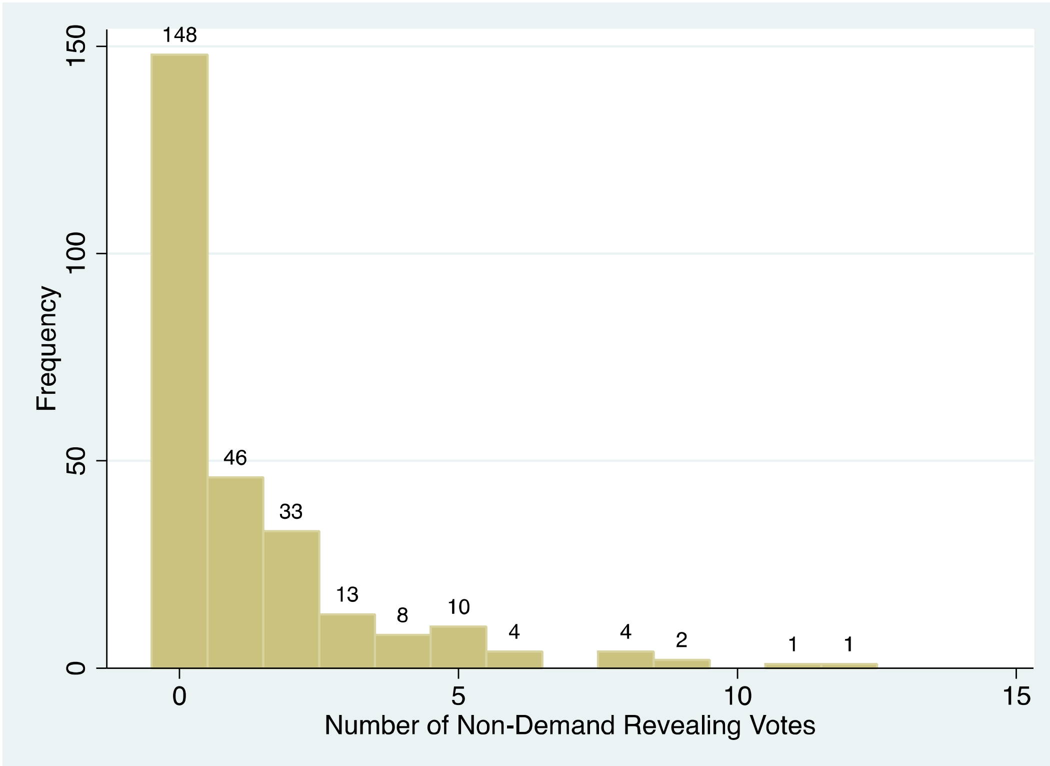

Figure 6. Frequency of number of non-demand revealing votes per subject (out of a possible 12)

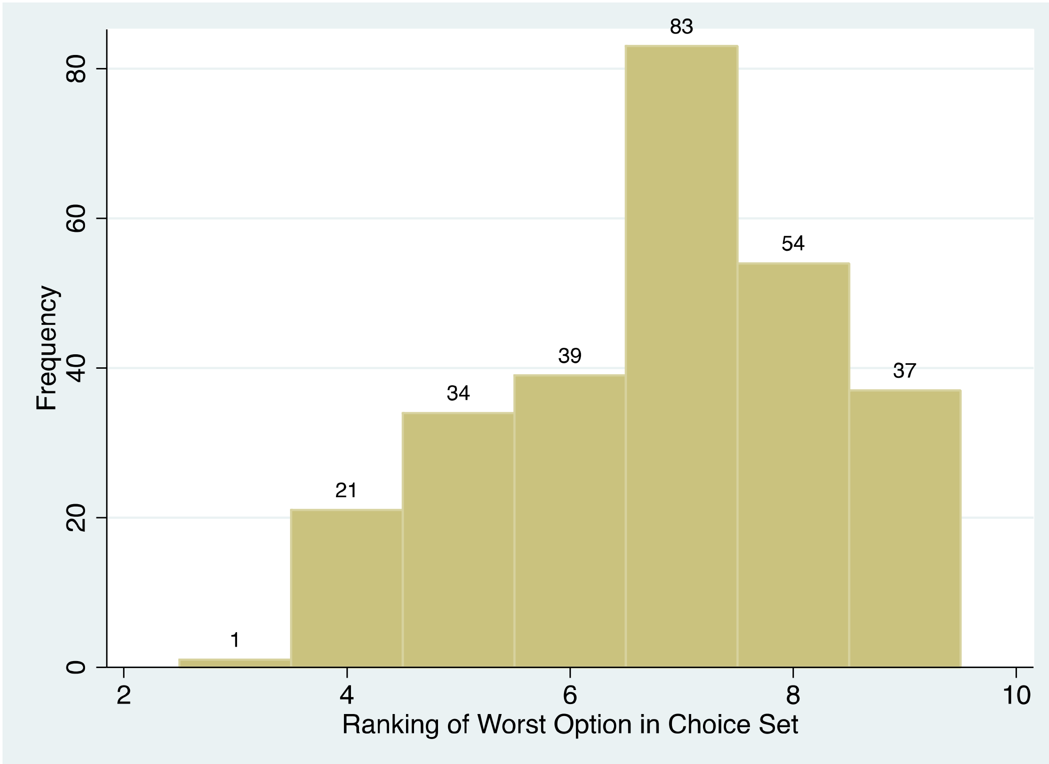

Figure 7. Ranking (out of 9) of the worst option in choice sets when a non-demand revealing vote for the second best option was cast

Table 1. Non-Demand Revealing (NDR) Choices in Experiment 1

Treatment	**Overall Fraction of Subjects Making NDR Choices**	**Fraction of NDR Choices for Status Quo**	**Fraction of NDR Choices for Other than Status Quo**
OSB	7%	4%	3%
OSM	8%	4%	4%
RB	4%	3%	1%
RM	11%	9%	2%

H0: Rate of NDR in Treatment A = Rate of NDR in Treatment B
(T-test of proportions with standard errors clustered at experimental session level)

	OSB	RM
OSM	p = 0.378	p = 0.291
RB	p = 0.223	p = 0.027
RM	p = 0.089	

Table 2. Percent of Non-Demand Revealing Choices by Induced Values:

Repeated Multiple Choice Treatment – Experiment 1

Option 1 Earnings	**Option 2 Earnings**	**Status Quo Earnings**	**% NDR Status Quo Choices**	**% NDR Non-Status Quo Choices**	**% of NDR Choices for 2nd Best Option**	**% of NDR Choices for 3rd Best Option**
\$5	\$14	\$10	14%	0%	100%	0%
\$3	\$5	\$10	0%	6%	50%	50%
\$5	\$17	\$10	11%	0%	100%	0%
\$19	\$8	\$10	14%	0%	100%	0%
\$19	\$3	\$10	6%	3%	67%	33%
\$17	\$19	\$10	3%	11%	80%	20%
\$14	\$8	\$10	8%	0%	100%	0%
\$14	\$3	\$10	17%	0%	100%	0%
\$8	\$17	\$10	8%	0%	100%	0%
Overall			9%	2%	92%	8%

Table 3. Proportion of Choice Sets with Status Quo as First or Second-Best Option when the Fraction of Choice Options Preferred to the Status Quo (π) is 0.50

		Number of Attributes (m)		
Number of Levels of each Attribute		2	3	4
	2	0.833	0.786	0.767
	3	0.781	0.760	0.753
	4	0.767	0.754	0.751

Table 4. Beliefs over Shares of Respondents Who Most Prefer Each Option for Orthogonal on the Difference Choice Sets

Choice Occasion	Share who Prefer Option 1	Share Who Prefer Option 2	Share Who Prefer Status Quo
1	0%	57%	43%
2	86%	0%	14%
3	86%	0%	14%
4	0%	57%	43%
5	71%	0%	29%
6	14%	29%	57%
7	0%	86%	14%
8	29%	14%	57%
9	0%	71%	29%
10	29%	14%	57%
11	0%	71%	29%
12	57%	0%	43%

Table 5. Predicted and Actual Fraction of Choice Sets in Which the Status Quo is Second-Best or First- or Second-Best

	Fraction of Choice Sets in Which the Status Quo is Second-Best or First- or Second-Best		
	$\pi = 0.125$	$\pi = 0.50$	$\pi = 0.75$
Status Quo is 2^{nd} Best			
Full Factorial Prediction: n = *2*, m = *3*	0.25	0.57	0.43
Full Factorial Prediction: Limiting Case	0.22	0.50	0.38
Actual OOD	0.25	1.00	0.50
Actual UBAL	0.08	0.33	0.33
Actual RAND	0.17	0.67	0.42
Status Quo is 1^{st} or 2^{nd} Best			
Full Factorial Prediction: $n = 2$, $m = 3$	1.00	0.79	0.46
Full Factorial Prediction: Limiting Case	0.98	0.75	0.44
Actual OOD	1.00	1.00	0.50
Actual UBAL	1.00	0.75	0.42
Actual RAND	1.00	0.75	0.50

Table 6. Fraction of Non-Demand Revealing Votes for the Second-Best and Worst Options

	Mean	Robust Std. Err.	95% Conf. Int.
2nd Best: By Method to Create Fractional Factorial Choice Set Designs			
OOD	0.79	0.051	[0.69, 0.89]
UBAL	0.82	0.050	[0.72, 0.92]
RAND	0.83	0.046	[0.74, 0.92]
2nd Best: By Fraction of Choice Options Preferred to the Status Quo (π)			
$\pi = 0.125$	0.87	0.035	[0.80, 0.94]
$\pi = 0.50$	0.79	0.048	[0.70, 0.89]
$\pi = 0.75$	0.73	0.064	[0.61, 0.86]
Worst: By Method to Create Fractional Factorial Choice Set Designs			
OOD	0.21	0.051	[0.11, 0.31]
UBAL	0.18	0.050	[0.08, 0.28]
RAND	0.17	0.046	[0.08, 0.26]
Worst: By Fraction of Choice Options Preferred to the Status Quo (π)			
$\pi = 0.125$	0.13	0.035	[0.06, 0.20]
$\pi = 0.50$	0.21	0.048	[0.11, 0.30]
$\pi = 0.75$	0.27	0.064	[0.14, 0.39]

Table 7. Predicted and Observed Fractions of Second-Best Votes for the Status Quo Option

	Fraction of Second-Best Votes for the Status Quo		
	$\pi = 0.125$	$\pi = 0.50$	$\pi = 0.75$
Predicted by Mathematical Models and Experimental Choice Set Designs			
Full Factorial Prediction: n = 2, m = 3	0.25	0.57	0.43
Full Factorial Prediction: Limiting Case	0.22	0.50	0.38
Predicted by the OOD Fractional Factorial Choice Set Design	0.25	1.00	0.50
Predicted by the UBAL Fractional Factorial Choice Set Design	0.08	0.33	0.33
Predicted by the RAND Fractional Factorial Choice Set Design	0.17	0.67	0.42
Experimentally Observed Fraction of Second-Best Votes for the Status Quo			
Observed: OOD Experimental Results	0.03	1.00	0.40
(Robust SE)	(0.04)	(N/A)	(0.15)
Observed: UBAL Experimental Results	0.04	0.07	0.47
(Robust SE)	(0.03)	(0.03)	(0.18)
Observed: RAND Experimental Results	0.02	0.80	0.39
(Robust SE)	(0.02)	(0.16)	(0.10)

Table 8. Rates of Demand Revelation by Category

	Mean	Robust Std. Err.	95% Conf. Int.
By Method to Create Fractional Factorial Choice Set Designs			
OOD	0.93	0.015	[0.90, 0.95]
UBAL	0.87	0.019	[0.83, 0.91]
RAND	0.90	0.019	[0.86, 0.93]
By Fraction of Choice Options Preferred to the Status Quo (π)			
$\pi = 0.125$	0.85	0.024	[0.81, 0.90]
$\pi = 0.50$	0.92	0.015	[0.89, 0.95]
$\pi = 0.75$	0.92	0.016	[0.89, 0.95]

Supplemental Materials

1. Experiment 1 Design Details

Treatments

One Shot Binary (OSB) – subjects face a single choice occasion on which they choose between two options, one of which is the status quo option. (8 groups of 9 subjects, 72 subjects total).
One Shot Multiple (OSM) – subjects face a single choice occasion on which they choose among three options, one of which is the status quo option. (8 groups of 9 subjects, 72 subjects total).
Repeated Binary (RB) – subjects make a series of six choices. On each choice occasion, subjects choose between two options, one of which is the status quo option. (4 groups of 9 subjects, 36 subject total).
Repeated Multiple (RM) – subjects make a series of nine choices. On each choice occasion, subjects choose among three options, one of which is the status quo option. (4 groups of 9 subjects, 36 subjects total).

Rules of the Game

In the one-shot treatments, each subject is randomly assigned a ballot containing one choice set. Subjects cast their vote for one of the options on the ballot, and turn in their ballots to the moderator. The moderator counts the votes and announces the winning option. Ties are resolved via a coin toss (two-way tie) or the roll of a 6-sided die (three-way tie). Subjects calculate their earnings from the winning option using the induced value function described below, and receive their earnings in cash at the conclusion of the experiment.

In the repeated choice treatments, each subject receives a package containing a series of six (binary choice treatment) or nine (multiple choice treatment) ballots, in random order. On each ballot, subjects vote for one of the options. The moderator collects each ballot after subjects cast their votes. The binding ballot is determined by rolling a die. The moderator counts the votes on the binding ballot, and announces the winning option. Ties are resolved as in the one-shot treatments. Subjects calculate their earnings from the winning option on the binding ballot, and receive their earnings in cash at the conclusion of the experiment.

Induced Values

Subject earnings are calculated using the following induced value function:

$$Value\ of\ Winning\ Option = (3\ \times Units\ of\ A) + (7\ \times Units\ of\ B) - Cost$$

Units of A, units of B, and cost for the winning option may vary by subject, depending on the outcome of the random assignment of ballots.

Choice Set Design

The experiment employs a 2^3 choice set design. Each option in the choice set has two attributes, termed attribute A and attribute B, and a cost. Each attribute and the cost have two possible levels. The possible levels of attribute A are 1 and 2. The possible levels of attribute B are 1 and

3. The possible levels of the cost are 5 and 10. The status quo option contains one unit of attribute A and one unit of attribute B, at zero cost.
The 2^3 design of the choice sets results in eight choice options which can be combined into choice sets. Consistent with practice in many field studies, we remove the option that provides the most attributes at the least cost, and the fewest attributes at the highest cost, as little information about preferences may be derived from having subjects make these trivial choices. The remaining six choice sets are combined with the status quo to create the binary choice sets. There are 56 possible combinations of two unique choice options and the status quo into multiple choice sets. We eliminate choice sets containing trivial choices, and choice sets in which the two options are the same but the order is reversed, to arrive at nine possible choice sets for the multiple choice experiment. Table S1 reports the binary and multiple choice sets for experiment 1.

Subject Number and Average Earnings
Power analyses and results from previous experiments show that for the one-shot treatments, a sample of 72 subjects results enables one to correctly reject the null hypothesis that the predicted and observed vote distributions are the same about 80% of the time, using an alpha of 0.05 and relatively conservative assumptions about deviations from predicted votes. The power approaches one rapidly as the observed vote distributions deviation further from the theoretical prediction. Each of the one-shot treatments includes 8 experimental sessions with 9 subjects per session. Each of the repeated choice treatments includes 4 experimental sessions with 9 subjects per session. As the unit of observation is the choice occasion, this results in 216 observations in the repeated binary choice treatment, and 324 observations in the repeated multiple choice treatment. Oversampling in these treatments is necessary to ensure that the results are not driven by behavior in a single experimental session.

All subjects were students at an undergraduate institution in the United States. Average subject earnings for a 40-45 minute experimental session. As noted in note 5, given then average weekly earnings of the subject pool, there is no question that the earnings amount was salient to the subjects.

2. Full Derivation of Position of Status Quo in Full Factorial Choice Sets

Consider balanced choice sets, that is, choice sets in which every attribute has the same number of levels.

Let:
m = the number of attributes of each option in the choice set
n = the number of levels of each attribute

A full factorial design results in n^m possible choice options to be combined into choice sets.

There are $n^m(n^m-1)$ possible choice sets that consist of two distinct options and the status quo.

Let:

π = the proportion of choice options that are preferred to the status quo where $0 \leq p \leq 1$. $(1 - \pi)$ is then the proportion of choice options that are not preferred to the status quo (We can assume no indifference without loss of generality.)

Then:
πn^m choice options are preferred to the status quo
$(1 - \pi)n^m$ choice options are not preferred to the status quo

In choice sets consisting of two options plus the status quo:
$(\pi n^m)(\pi n^m - 1)$ will have the status quo as the worst option in the choice set
$[(1 - \pi)n^m][(1 - \pi)n^m - 1]$ will have the status quo as the best option in the choice set

Leaving

$$n^m(n^m - 1) - \{\pi n^m(\pi n^m - 1) + [(1-\pi)n^m][(1-\pi)n^m - 1]\}$$

Choice sets that have the status quo as the second best option.

The choice sets constitute the following fraction of the total number of choice sets:

$$\frac{n^m(n^m - 1) - \{\pi n^m(\pi n^m - 1) + [(1-\pi)n^m][(1-\pi)n^m - 1]\}}{n^m(n^m - 1)}$$

This fraction reduces to:

$$\frac{n^m[2\pi(1-\pi)]}{n^m - 1}$$

And the limit as either $n \to \infty$ or $m \to \infty$ is $2\pi(1-\pi)$.

If we add back in the $[(1 - \pi)n^m][(1 - \pi)n^m - 1]$ choice sets in which the status quo is the best, the fraction of choice sets in which the status quo is ranked either first or second is:

$$\frac{n^{2m}2\pi(1-\pi) + [(1-\pi)n^m][(1-\pi)n^m - 1]}{n^m(n^m - 1)}$$

This fraction reduces to:

$$\frac{(1-\pi)[n^m(1+\pi) - 1]}{n^m - 1}$$

And the limit of this fraction as either $n \to \infty$ or $m \to \infty$ is $(1-\pi)(1+\pi) = (1-\pi^2)$.

Table S1. Experiment 1 Choice Sets

Binary Choice Sets											
Option 1				Option 2				Status Quo			
A	B	Cost	Earnings					A	B	Cost	Earnings
1	1	5	$5	N/A				1	1	0	$10
1	3	5	$19	N/A				1	1	0	$10
1	3	10	$14	N/A				1	1	0	$10
2	1	5	$8	N/A				1	1	0	$10
2	1	10	$3	N/A				1	1	0	$10
2	3	10	$17	N/A				1	1	0	$10
Multiple Choice Sets											
Option 1				Option 2				Status Quo			
A	B	Cost	Earnings	A	B	Cost	Payoff	A	B	Cost	Earnings
1	1	5	$5	1	3	10	$14	1	1	0	$10
2	1	10	$3	1	1	5	$5	1	1	0	$10
1	1	5	$5	2	3	10	$17	1	1	0	$10
1	3	5	$19	2	1	5	$8	1	1	0	$10
1	3	5	$19	2	1	10	$3	1	1	0	$10
2	3	10	$17	1	3	5	$19	1	1	0	$10
1	3	10	$14	2	1	5	$8	1	1	0	$10
1	3	10	$14	2	1	10	$3	1	1	0	$10
2	1	5	$8	2	3	10	$17	1	1	0	$10

Table S2. Rate of Non-Demand Revelation (NDR) by Choice Occasion: Experiment 1

	Overall Fraction of Subjects Making NDR Choices		Fraction of NDR Choices for Status Quo		Fraction of NDR Choices for Other than Status Quo	
	Binary	Multiple	Binary	Multiple	Binary	Multiple
Choice #1	6%	6%	100%	100%	0%	0%
Choice #2	0%	19%	0%	71%	0%	29%
Choice #3	3%	11%	100%	100%	0%	0%
Choice #4	8%	8%	67%	100%	33%	0%
Choice #5	3%	11%	100%	100%	0%	0%
Choice #6	6%	14%	50%	60%	50%	40%
Choice #7	N/A	6%	N/A	50%	N/A	50%
Choice #8	N/A	11%	N/A	100%	N/A	0%
Choice #9	N/A	14%	N/A	60%	N/A	40%

Table S3. Values of Pi and Marginal Values of Attributes in Choice Set Design Simulations

Pi	Marginal Value of Attribute A	Marginal Value of Attribute B
0	2	1
0.125	3	2
0.250	2	3
0.375	3	5
0.500	3	7
0.625	6	7
0.750	11	7

Table S4. Orthogonal on the Differences (OOD) Choice Sets

	Option 1			Option 2		
Choice Occasion	Units of Attribute A	Units of Attribute B	Cost	Units of Attribute A	Units of Attribute B	Cost
1	1	1	5	2	3	10
2	2	3	5	1	1	10
3	2	3	5	1	1	10
4	1	1	5	2	3	10
5	1	3	5	2	1	10
6	2	1	5	1	3	10
7	1	1	10	2	3	5
8	1	3	10	2	1	5
9	2	1	10	1	3	5
10	1	3	10	2	1	5
11	2	1	10	1	3	5
12	2	3	10	1	1	5

Table S5. Utility Balanced Choice Sets, $\pi = 0.125$

	Option 1			Option 2		
Choice Occasion	Units of Attribute A	Units of Attribute B	Cost	Units of Attribute A	Units of Attribute B	Cost
1	1	1	5	2	1	5
2	2	1	5	1	3	10
3	2	3	10	1	3	5
4	1	3	5	2	3	10
5	1	3	5	2	1	5
6	2	3	5	2	1	5
7	1	1	10	1	3	10
8	1	1	10	2	1	5
9	2	1	5	1	3	5
10	1	1	5	1	3	10
11	2	3	10	2	1	5
12	2	1	5	2	3	10

Table S6. Utility Balanced Choice Sets, $\pi = 0.50$

	Option 1			Option 2		
Choice Occasion	Units of Attribute A	Units of Attribute B	Cost	Units of Attribute A	Units of Attribute B	Cost
1	2	3	10	1	3	5
2	2	1	10	2	1	5
3	2	1	10	1	3	10
4	2	1	5	1	1	5
5	2	1	5	2	1	10
6	1	3	10	2	1	5
7	1	3	10	1	1	10
8	1	1	5	2	1	5
9	1	3	5	1	3	10
10	2	1	10	1	1	10
11	2	1	5	2	3	10
12	1	3	10	2	3	10

Table S7. Utility Balanced Choice Sets, $\pi = 0.75$

	Option 1			Option 2		
Choice Occasion	Units of Attribute A	Units of Attribute B	Cost	Units of Attribute A	Units of Attribute B	Cost
1	1	3	10	2	3	10
2	1	3	10	2	1	10
3	2	3	5	1	3	5
4	2	1	5	2	3	10
5	1	1	5	1	3	10
6	1	1	10	1	1	5
7	1	1	5	2	1	10
8	2	1	5	1	3	10
9	1	3	10	2	1	5
10	2	1	10	1	3	10
11	2	3	5	1	1	10
12	1	3	10	1	1	5

Table S8. Randomly-Generated Choice Sets

	Option 1			Option 2		
Choice Occasion	Units of Attribute A	Units of Attribute B	Cost	Units of Attribute A	Units of Attribute B	Cost
1	2	3	5	1	3	5
2	2	3	5	2	1	5
3	2	1	10	1	3	10
4	1	3	10	1	1	10
5	1	3	5	2	1	5
6	1	1	5	1	3	5 or 10*
7	1	1	5	1	3	5
8	1	3	10	2	3	10
9	2	3	10	1	3	10
10	1	1	10	2	3	10
11	1	1	5	1	1	5
12	2	3	10	1	1	10

*For $\pi = 0.125$, Option 2 contained a Cost of 5. For all other subjects, cost was 10.

Table S9. Subject Demographic Characteristics

	Number	Percent
Male	209	77.4%
Female	50	18.5%
Declined to Answer	11	4.1%
Freshman	37	13.7%
Sophomore	134	49.6%
Junior	38	14.1%
Senior	51	18.9%
Declined to Answer	10	3.7%
Technical Major	117	43.3%
Non-Technical Major	113	41.9%
Major Undeclared	30	11.1%
Declined to Answer	10	3.7%

Table S10. Rates of Demand Revelation by Subject Demographics

	Mean	Robust Std. Err.	95% Conf. Int.
By Gender			
Male	0.91	0.011	[0.89, 0.93]
Female	0.88	0.023	[0.83, 0.92]
Unknown	0.79	0.085	[0.62, 0.96]
By Class Year			
Freshman	0.91	0.026	[0.86, 0.96]
Sophomore	0.88	0.015	[0.85, 0.91]
Junior	0.92	0.020	[0.88, 0.96]
Senior	0.94	0.017	[0.91, 0.97]
Unknown	0.77	0.091	[0.59, 0.95]
By Major Field of Study			
Technical	0.90	0.017	[0.86, 0.93]
Non-Technical	0.91	0.012	[0.89, 0.93]
Undeclared	0.89	0.032	[0.83, 0.95]
Unknown	0.77	0.091	[0.59, 0.95]