January 31, 2010

Subjective choice difficulty in stated choice tasks

Eric Duquette, Trudy Ann Cameron*
Department of Economics, 435 PLC, University of Oregon
Eugene, OR 97403-1285

and

J.R. DeShazo
Department of Public Policy and Institute of the Environment, UCLA
Los Angeles, CA 90095-1656

*Corresponding author: cameron@uoregon.edu

Keywords: random utility model, stated preference, choice difficulty, choice complexity, context effects, ancillary conditions

JEL Classifications: C42, C35, Q51, I18

Acknowledgements: This research has been supported in part by a grant from the National Science Foundation (SES-0551009) to the University of Oregon (PI: Trudy Ann Cameron). It employs original survey data from an earlier project supported by the US Environmental Protection Agency (R829485) and Health Canada (Contract H5431-010041/001/SS). This project was also supported in part by resources from the Raymond F. Mikesell Foundation at the University of Oregon. This work has not been formally reviewed by any of the sponsoring agencies. Any remaining errors are our own.

Subjective choice difficulty in stated choice tasks

Abstract

The existing literature has addressed choice set complexity, as well as other choice set characteristics, as ancillary conditions (i.e. context effects) that can affect performance in stated choice tasks. Choice set characteristics can contribute to the perceived difficulty of a choice task and can lead respondents to adopt choice heuristics which may suggest that they are not choosing rationally based upon the full complement of information provided in the choice scenario. However, "choice difficulty" is not usually observable by the researcher. Objective measures of choice set complexity have instead been assumed to proxy for choice difficulty and these measures have been used empirically to shift the scale of the error term or the slope coefficients in choice models. In our stated preference survey, respondents are asked directly to rate the subjective difficulty of each of their choices. We use this unique opportunity to explore the determinants of subjective choice difficulty to assess how well the customary reduced-form proxies are likely to capture this behavioral aspect of subjects' interactions with choice tasks. Common measures do not fully explain subjective choice difficulty, which also depends on the interplay among objective attribute-space complexity, the similarity of alternatives in utility space, and cognitive resource constraints.

Keywords: random utility model, stated preference, choice difficulty, choice complexity, context effects, ancillary conditions

JEL Classifications: C42, C35, Q51, I18

Subjective choice difficulty in stated choice tasks

1. Introduction

The stated preference literature abounds with examples of consumer choice data from valuation surveys that probe individuals' preferences for non-market or pre-test-market goods. To properly estimate preferences and derive willingness to pay (*WTP*) measures, individual-level choice data require empirical models that can handle heterogeneity in preferences, in decision heuristics, and in the context variables that constitute the "ancillary conditions" of the choice environment and contribute to an individual's overall preferred bundle of goods (as in Bernheim and Rangel (2009)). For empirical choice models based in random utility theory, it is unclear exactly how these models should include other aspects of choice that are apart from the formal minimalist representation of a utility function that individuals are typically assumed to maximize.

In this paper, we emphasize the notion of "choice difficulty" as an aspect of the behavioral context for a choice task. If ignored by the researcher, choice difficulty can lead to apparent inconsistencies in the outcomes of utility maximization, even after conditioning on observable determinants of an individual's preferences and the salient features of a good. We elicit subjective choice difficulties and explore their determinants. Our goal is to evaluate the potential role of explicit subjective choice difficulty measures as important adjuncts to choice modeling, especially since they have the potential to index more comprehensively the variety of choice set features that are more typically employed, in reduced form specifications, as controls for unobserved subjective choice difficulty.

The typical strategy used to investigate context effects in stated preference research has been to assess the influence of different types of survey design elements (such as the number of attributes, alternatives, or choice tasks) on individuals' decision-making behavior. Existing studies commonly refer to these types of objectively measured external influences stemming from the survey design as "choice complexity" and have shown that various dimensions of choice complexity can significantly impact estimates of the marginal utility parameters and thus the resulting calculations of WTP (see Louviere, et al. (2002), Louviere, et al. (2005), and Adamowicz and DeShazo (2006)).[1]

We extend the usual definition of choice complexity to build a broader notion of "choice difficulty." Choice difficulty encompasses interactions between choice set complexity and respondent characteristics, such as sociodemographic traits, idiosyncratic subjective experiences, cognitive capacity, interest in the task, and current attention budgets. Our direct measure of choice difficulty, unique to this survey, comes from a follow-up question that elicits each individual's subjective assessment of the difficulty of the conjoint choice task just completed.[2] We show that our respondents' subjective choice difficulty ratings are correlated with (1) some common objective measures of choice set complexity, (2) observed individual characteristics that may proxy for better abilities (or opportunities) to make consistent choices, and (3) variables based upon evidence from elsewhere in the same survey, for that respondent, that may capture other factors expected to influence the perceived difficulty level for the choice task in question.

[1] In a revealed-preference setting, Beshears, et al. (2008) discuss factors that can potentially contribute to decision-making errors and, thus, a disparity between revealed preferences and "normative preferences"—preferences that represent an individual's true interests. They identify five important factors that contribute to the disparity. These include passive choice, complexity, limited personal experience, third-party marketing, and intertemporal choices.

[2] The elicitation of respondents' subjective impressions about their earlier survey responses has also been used in the literature on preference uncertainty e.g.Evans, et al. (2003), Li and Mattsson (1995), Vossler, et al. (2003), and Welsh and Poe (1998). In this literature, researchers incorporate subjective measures of preference *uncertainty* into the estimation process to improve *WTP* estimates that might otherwise be biased. In our analysis, we recognize that the effects of preference uncertainty and choice difficulty on choice outcomes are likely to be correlated. As our results indicate, cognitive capacity can play a large role in choice difficulty, and this is also likely true for preference uncertainty. However, choice difficulty can arise even when respondents are certain about their preferences, as we explain in detail in Section 2.

Our research also explores an additional candidate measure for choice difficulty that quantifies the distance between alternatives in *utility space*, rather than *attribute space*. With the exception of the entropy measure employed by Swait and Adamowicz (2001a, 2001b), all other common empirical measures of choice complexity are, conceptually, distances between choice-set alternatives in *attribute* space. These attribute-space measures can be problematic in that they can fail to capture the type of choice complexity that arises when alternatives are far apart in attribute-space but nevertheless close in terms of a particular individual's utility function (i.e. when two alternatives are far apart in attribute-space but still lie close to the same indifference curve).[3] In comparison to entropy, our utility-space measure of the similarity of alternatives can reflect the same types of heterogeneity, but may be easier to interpret. We find that, like entropy, our measure is strongly correlated with subjective choice difficulty and its effects on respondents' ratings of choice difficulty are consistent with our priors. However, neither our alternative measure, nor entropy, is the *only* systematic determinant of subjective choice difficulty.

In contrast to previous studies, we do not simply embed proxies for choice difficulty directly into our conjoint choice model, using these proxies to shift either the estimated utility parameters or the scale factor (error dispersion) for the choice model. Instead, the attributes of each choice set, in some cases along with the estimated utility parameters from a preliminary choice model, are used to build an array of variables which we use to analyze the determinants of perceived choice difficulty. Among the proxies normally used to control for choice difficulty, we explore which candidates seem to do the best job. Specifically, we investigate the factors which may contribute additional explanatory power. Furthermore, if the subjective difficulty variable

[3] Utility-space measures also have the potential to be unique for each individual when preferences are allowed to be heterogeneous.

adequately captures the various proxy variables which have been used elsewhere in the literature then it might be a good practice to attempt to elicit subjective choice difficulties directly in all stated preference surveys. Our results suggest that this is a strong possibility—the fitted values from an estimated choice difficulty model could function as a single-valued index variable. This index could be used to purge a fitted choice model of systematic variation in marginal utility parameters (or implied *WTP*) due to choice sets which are outliers in terms of choice difficulty, for at least some respondents, without vastly increasing the size of the parameter space of the model.

The rest of this paper proceeds as follows. In Section 2, we describe choice complexity measures considered in the existing literature and introduce additional measures that may explain the subjective choice difficulty of respondents. In Section 3, we provide a brief overview of the survey methodology and describe our stated choice data along with other individual-specific characteristics that we have for our respondents. Section 4 describes the basic indirect utility specification that we use to model individuals' responses to the conjoint survey so that we can build measures of the closeness of alternatives in utility space. This section also outlines the empirical specifications for modeling subjective choice difficulty. We present our key findings in Section 5 and discuss some possible extensions in Section 6. Section 7 concludes.

2. Dimensions of Choice Difficulty

We entertain a variety of different dimensions of choice complexity which may contribute to a respondent's perception that a choice task is more or less difficult. We consider measures which are calculated in utility space and therefore require preliminary estimation of a preference function before complexity can be quantified. We also consider an array of complexity measures calculated simply in attribute space, which are independent of the preferences of the individual who is making the choice. Finally we will assess the impact on subjective choice difficulty of some available variables that may reflect cognitive capacity or cognitive constraints, plus a variety of individual respondent characteristics and even some variables that describe observed patterns of respondent behavior across the five choice tasks that most respondents completed.

2.1 Objective Measures of Choice Complexity in Utility Space

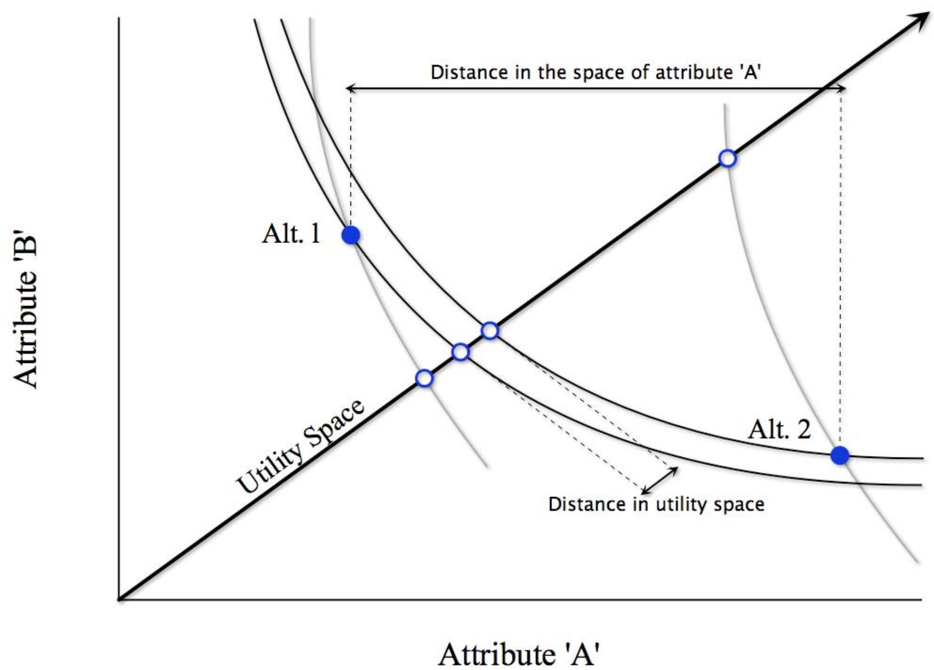
The most typical measures of choice set complexity are based upon calculations in attribute space. The information contained within these measures may encompass the nature of the different alternatives (goods) in the choice set, their number, the level of detail used (or needed) to describe them, and the number of choice sets presented to the individual. However, we suspect that important aspects of choice complexity can also originate from individual *preferences* over the different attributes describing each alternative, in combination with the *levels* of the attributes themselves.

In the quantification of objective choice set complexity, Swait and Adamowicz (2001a, b) propose a measure known as entropy. This measure was first introduced in the field of *information theory* by Shannon (1948) and can be defined over any set of probabilistic events. Probabilistic events with relatively large degrees of uncertainty will have outcomes that reveal relatively greater amounts of information, or entropy. Swait and Adamowicz (2001a, b) define the entropy of a choice set to be a function of the choice probabilities associated with each alternative in a choice set. Random-utility theory assigns a choice probability to each alternative, $\pi_{ij} = \pi\left(x_{ij}\hat{\beta}\right)$, that is a function of the latent utility associated with each alternative as measured by the estimated value of the utility index, $x_{ij}\hat{\beta}$, for a conditional logit choice model. Thus, the entropy of a choice set, $H_i = -\sum_{j=1}^{J} \pi_{ij} \log\left(\pi_{ij}\right)$, can potentially capture choice complexity stemming both from the levels of attributes, x_{ij}, and preferences over those attributes, β.

Entropy is minimized when there is one dominant alternative in the choice set, and maximized if each alternative is equally likely. The authors model the “scale factor” in a random utility model (the inverse of the error variance) as a quadratic function of entropy and are able to identify systematic effects on choice consistency related to both the linear and the squared terms in entropy. Swait and Adamowicz (2001a) find that the estimated variance (noise) for the utility function is first increasing, and then decreasing, in the level of entropy for a choice set.

The most important thing about the entropy measure is that it incorporates utility-space information, whereas most other existing measures of complexity are restricted simply to attribute space and are assumed to be independent of individual preferences. In the two-alternative case, Figure 1 depicts the relationship between proximity of alternatives in attribute space and proximity in utility space. These two types of proximity can potentially have independent effects on the difficulty of a choice task. In Figure 1, after Mas-Colell, et al. (1995), the utility of an alternative is represented by the intersection of the utility curve in attribute space with the 45° line. The standard deviation of utility is a measure of the “distance” among alternatives in a one-dimensional utility space, as opposed to a multi-dimensional attribute space. As illustrated, the utility-space distance of a set of alternatives can vary independently from attribute-space distance, for example, as measured by the variability of any single attribute’s levels across alternatives. The steeper (lighter) utility curves represent the possible preferences of a different individual and suggest that preferences can significantly affect the amount of complexity measured in utility space, even for the same set of alternatives. Thus, any complexity

Figure 1 - Attribute- versus Utility-Space Complexity



.

that arises in utility-space is strongly dependent on an individual's preferences over the attributes of the alternatives. For more than two alternatives, of course, a simple distance measure alone is inadequate. To measure the extent to which there is a clear winner in utility space, entropy is a potentially more useful summary.

Entropy is thus one way to very succinctly encapsulate choice set complexity in a way that also reflects individual preferences. Swait and Adamowicz (2001a) explicitly invoke the contribution of this complexity to *choice difficulty* in formulating their hypotheses regarding complexity and variance (p. 158):

> "Because complexity is hypothesized to demand additional outlays of effort on the part of consumers to find the utility-maximizing choice, we expect that variance (scale) will be increasing (decreasing) in complexity."

Since they have no direct measure of choice difficulty, Swait and Adamowicz must simply *assume* that choice difficulty is the unobserved behavioral link between entropy as a convenient one-dimensional summary of choice complexity and the resulting observed heteroscedasticity in choice models. In this paper, we use our direct measure of choice difficulty to assess the extent to which subjective choice difficulty is related to choice set entropy, other candidate measures of complexity, measures of cognitive capacity and/or constraints, sociodemographic/attitudinal variables, and even some systematic patterns in observed choice behaviors.

Along with the entropy measure proposed by Swait and Adamowicz (2001a, b), we consider an alternative and somewhat simpler utility-space measure: the standard deviation of the systematic component of the estimated utility across alternatives in the choice set. In contrast to entropy, this alternative measure is simply a summary of the extent to which estimated utility-levels differ across alternatives for each individual. When the utility-differences across alternatives in the choice set are smaller, an individual's most-preferred alternative will be more difficult to discern. Entropy does a better job of identifying large positive outliers in terms of utility, but it unavoidably subsumes an additional dimension—the number of alternatives in the choice set. This is moot when all choices involve the same number of alternatives, but may be relevant in cases where the sizes of choice sets vary.[4]

2.2 Objective Measures of Choice Complexity in Attribute Space

Aside from the entropy measure, most researchers have adopted the term "choice complexity" to describe the systematic influence of survey design elements and/or the design of the individual choice sets on response patterns, independent of the characteristics of any particular respondent. DeShazo and Fermo (2002) demonstrate the effects of choice complexity on "choice consistency" via the scale of the error term in a random utility model. They find choice consistency to be systematically affected by the number of alternatives, the number of attributes per alternative, and the number of attributes which are constant across alternatives. In addition, they calculate a measure of the standard deviation of attribute levels within each alternative, and then compute the across-alternative mean of these standard deviations, as well as the across-alternative standard deviation of these within-alternative standard deviations.[5] These additional choice set properties, collectively described as the "information structure" for each choice set, are also shown to have systematic effects on the consistency of responses by individuals. Hensher and his co-authors (see Hensher (2004, 2006a, b)) likewise include a three-

[4] Yet another measure of choice difficulty in utility space might be the size of the "lead" held by the highest-utility alternative.

[5] Each attribute in the DeShazo and Fermo study is offered at one of just three equally spaced levels (which can be denoted as -1, 0, or +1). This finesses the problem of different units of measurement for each alternative, but it may confound interpretation of the standard deviation across attribute levels because of the inherent scale differences across attributes. This may compromise the estimated effects of the complexity measures (see Lancsar, et al. (2007)). In contrast, but analogously, the attributes we use are cardinal measures so that they have well-defined scales of measurement.

level attribute-space measure (wide, narrow, and base) for the range of each attribute level as an objective measure of complexity and find varying evidence for the effects of these measures on choice consistency.

For more general types of attributes, our analysis attempts to improve upon the purely objective measures of information structure used in DeShazo and Fermo (2002). Our attribute levels are cardinal variables. Prior to calculating the standard deviation of the levels of different types of attributes within an alternative, we first standardize the scales of measurement for each attribute (see Appendix B). Standardization prevents the different scales of measurement of the different attributes from acting like weights on their influence. If each alternative is good on all attributes, or undesirable on all attributes, the choice task can be expected to be easier. If, instead, each alternative is good on some dimensions and undesirable on others, respondents will have to make more types of tradeoffs among attributes to identify the most-preferred alternative. To measure these tendencies within a given choice set, we use the choice-set-level variables $Mean\ SD_{i\bullet k}$ and $Disp.\ SD_{i\bullet k}$ (as developed in Appendix B).

As do Hensher and his co-authors (see Hensher (2004, 2006a, b)) and Johnson (2006), we also employ continuous measures of the standard deviation, across alternatives in the choice set, for the levels of each attribute. We refer to these separate measures as descriptions of the "across-alternative attribute variability." The impacts of these components of complexity on choice difficulty are theoretically indeterminate. A large standard deviation of an attribute's level may place cognitive stress on a respondent by forcing him or her to actively consider regions of attribute space that may not be contained in the respondent's everyday choice set. Likewise, a small range could also be a source of stress if a respondent lacks the cognitive ability to discriminate between small differences in attribute levels. In the limit however, a small enough range could render the alternatives nearly indistinguishable along that dimension of the attribute in question, reducing the number of attributes remaining to be compared and thus making the choice task easier. Thus quadratic forms or other nonlinear relationships need to be explored.

2.3 Observable Individual Characteristics

In this paper, we consider the effects of observable respondent characteristics directly upon subjective choice difficulty. The existing choice literature also considers the effects of these types of variables on the decision making process, but only in a reduced-form sense—by using them as direct shifters of either the marginal utility parameters or the scale of the error term. For example, Hensher, et al. (2005), Hensher (2006a), and Hensher, et al. (2007) study the effects of choice set complexity using an elegant design-of-designs (DoD) approach.[6] Even though these studies, and others, consider income and age, there are likely to be many other individual characteristics that could indirectly affect the choice outcomes of respondents via their effects on subjective choice difficulty. We entertain a wide array of sociodemographic characteristics and other respondent-specific factors as potential covariates for choice difficulty.

2.4 Observable Measures of Cognitive Resource Constraints

We consider both educational attainment, and response times for other choice tasks by the same respondent, as objective proxies for cognitive resource constraints that potentially co-vary with choice difficulty. Educational attainment, in part, may reflect an individual's ability to make decisions under increasingly difficult choice scenarios. If so, greater educational attainment may lead to lower average subjective ratings of choice difficulty. Further, an

[6] Choice sets with different complexity characteristics are assigned randomly across split samples of respondents. In their mixed logit models, Hensher and his collaborators constrain to zero the marginal utilities associated with the attributes which each individual self-reports to have ignored in making their choices.

individual's capacity to process a difficult choice may affect the time required to make each choice. Thus, longer response times may also be associated with more difficult choices (although they may also belie looser time constraints). If a tighter time constraint results in shorter response times, then choices may be judged to be more difficult.

Several existing studies provide evidence to support the potential usefulness of these "cognitive" measures to explain choice behavior. Haaijer, et al. (2000) and Rose and Black (2006) and allow the scale factor (error variance), or the variances of slope coefficients in a random parameters choice model, to depend upon the response times (response latencies) of individuals. Both studies find large improvements in explanatory power over models which do not incorporate information on response times. In a non-economic social science choice context, Fischer, et al. (2000) show that the within-alternative "attribute conflict" that arises due to variation in the attribute levels of an alternative contributes to longer response times and noisier responses in choices among of alternatives.[7]

3. The Stated Preference Survey Data

3.1 The Survey Design

Our analysis uses an existing large sample of stated preference survey data concerning preferences with respect to privately supplied programs to reduce health risks. We also take advantage of the random-utility-based theoretical model developed in Cameron and DeShazo (2009) as a basic framework for our analysis.[8] This nationally representative survey includes

---

[7] In Fischer, et al. (2000), each respondent in an experimental setting provides a preference rating for a set of twenty alternatives that are presented sequentially and then ratings for an identical set of alternatives, but with a different randomized ordering, after a period of "filler." Therefore, respondents rate each of the assigned alternatives twice. The authors use the difference in rating for an alternative as a measure of response error.

[8] For more information on the survey instrument and the data, see the appendices which accompany Cameron and DeShazo (2009): Appendix A – Survey Design & Development, Appendix B – Stated Preference Quality Assurance and Quality Control Checks, Appendix C – Details of the Choice Set Design, Appendix D – The Knowledge

adults aged 25 years and older in the United States.[9] In brief, the stated preference survey consists of five modules. The first module asks respondents about their subjective risks of contracting the major illnesses or injuries which are the focus of the survey, the extent to which lifestyle changes might reduce their risks of these illnesses, and how taxing it might be to implement these lifestyle changes.

The second module is a tutorial that explains the concept of an "illness profile," which is a sequence of prospective future health states. An illness profile has attributes that include the number of years before the individual becomes sick (also referred to as the latency of the illness), illness-years while the individual is sick, recovered/remission years after the individual recovers from the illness, and lost life-years if the individual dies earlier than he would have without the disease or injury. Then the tutorial informs the individual that he might be able to purchase a new diagnostic testing program, at a monthly cost, that would reduce his risk of experiencing each illness profile.[10]

The third, and key, module of each survey consists of five different three-alternative conjoint choice experiments where the individual is asked to choose between two possible health-risk reduction programs and a status quo alternative. The survey design is essentially orthogonal, in that each illness program attribute—monthly program cost, risk reduction, the latency of the illness, its duration, and the lost-life years—is randomized across alternatives, choice occasions, and individuals. In addition, a "label" for each illness profile is randomly selected from five specific types of cancer, heart attack, heart disease, stroke, respiratory illness,

---

Networks Panel and Sample Selection Corrections, Appendix E – Model, Estimation and Alternative Analyses, and Appendix F – Estimating Sample Codebook.

[9] Knowledge Networks, Inc administered an internet survey to a sample of 2,439 of their panelists with a response rate of 79 percent.

[10] Each illness-related risk-reduction program consists of diagnostic blood tests, drug therapies, and life-style changes, the costs of which would need to be paid annually, since they would not be covered by health insurance.

diabetes, traffic accident or Alzheimer's disease (with occasional exclusions based on plausibility). Illness profiles need to be unique to each age/gender combination, so simple randomization proved more viable than any attempt at some type of fractional factorial design. One single example of a randomized choice scenario from the survey is presented in Figure 2.

Each choice exercise is immediately followed by a set of debriefing questions designed to help the researcher understand the individual's reasons for their particular choice. Some debriefing questions depend on the alternative chosen by the respondent—in particular, those who choose the status quo alternative ("Neither Program") are asked why it is their preferred alternative. Other debriefing questions, including the key "choice difficulty" question for this paper, are asked regardless of which alternative the individual selects. The crucial question for this paper, shown in Figure 3, is "How difficult was your choice on the previous screen?" Subjects were invited to respond on a Likert-type scale from 1="easy" to 7="very difficult."

The fourth module of the survey contains additional debriefing questions that permit us to explore other potential determinants of the individual's responses. A final module is collected separately from the same consumer panel and contains the respondent's socio-demographic characteristics and a detailed medical history, including which major diseases the individual has already faced.

3.2 Data Description

Our data set contains information on 1789 individuals who collectively made choices from a total of 8807 choice sets.[12] With these data, we are not able to study the effects of the number of alternatives and the number of attributes on subjective choice difficulty because these

[12] Of the 8817 choice sets with otherwise sufficiently complete data for analysis, 10 are dropped because each of these choice sets is the sole usable choice set for a respondent.

Choose the program that reduces the illness that you most want to avoid. But think carefully about whether the costs are too high for you. If both programs are too expensive, then choose Neither Program.

If you choose "neither program", remember that you could die early from a number of causes, including the ones described below.

	Program A for Diabetes	**Program B for Heart Attack**
Symptoms/ Treatment	Get sick when 77 years-old 6 weeks of hospitalization No surgery Moderate pain for 7 years	Get sick when 67 years-old No hospitalization No surgery Severe pain for a few hours
Recovery/ Life expectancy	Do not recover Die at 84 instead of 88	Do not recover Die suddenly at 67 instead of 88
Risk Reduction	10% From 10 in 1,000 to 9 in 1,000	10% From 40 in 1,000 to 36 in 1,000
Costs to you	$12 per month [= $144 per year]	$17 per month [= $204 per year]
Your choice	☐ **Reduce my chance of diabetes**	☐ **Reduce my chance of heart attack**

☐ **Neither Program**

Figure 2 – Example of a Choice Scenario

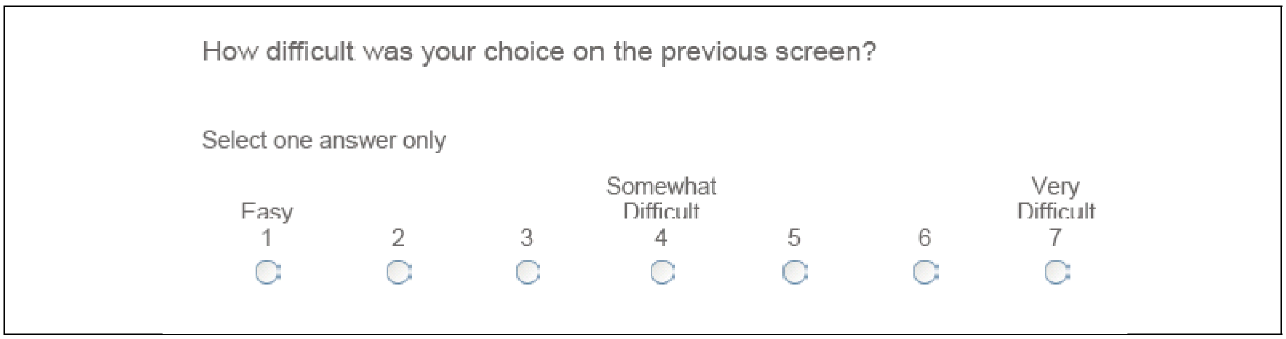


Figure 3 – Wording of the Follow-up Question Concerning Choice Difficulty

dimensions of the choice scenarios are held constant across all five choice sets posed to each respondent (i.e. all choice sets have three alternatives, and every alternative is described in terms of the same list of attributes). However, this restriction still allows us to study many of the other potential determinants of choice difficulty. Table 1 summarizes the key variables in our analysis.

How difficult, our dependent variable, consists of the 1-to-7 subjective difficulty rating by respondents for each choice, with higher-numbered categories conveying greater difficulty. Figure 4 displays the distribution of difficulty ratings by choice occasion. The average difficulty rating for individuals is 2.88. The remaining variables in Table 1 are potential determinants of choice difficulty. We divide these determinants into three broad categories: objective measures of choice set complexity, observable sociodemographic characteristics of the respondent, and proxies for the likely cognitive resources or constraints for each individual respondent.

As in previous studies which consider objective measures of choice set complexity, we employ a number of constructed variables. We consider Swait and Adamowicz's entropy value as one measure of complexity in utility space, but we also consider the simpler standard deviation of the fitted utility indices across alternatives, *Std. dev. of fitted U*, as an alternative. Figure 5 shows the distribution of *Std. dev. of fitted U* across choice occasions, based on the estimated utility parameters from a preliminary choice model that we discuss in the following section. Choice set attributes were random by construction, of course. Thus, the distribution of *Std. dev. of fitted U* should be unchanged across the five different choice occasions even though this measure is preference-dependent.

In addition to the two possibilities for utility-space measures of choice set complexity, we also examine some of the other customary measures within attribute space. Following DeShazo and Fermo (2002), these measures might include the mean and dispersion across alternatives of

Table 1 – Summary Statistics (n=22176)

	Mean	Std. dev.	Min.	Max.
Dependent variable				
How difficult (1 very easy - 7 very hard)	2.87	1.69	1	7
Measures of choice set complexity				
(a) In utility space:				
Std. dev. of fitted U	0.18	0.08	8e-4	0.48
Entropy	1.09	0.01	1.03	1.1
(b) In attribute space:				
(1) Within-alternative attrib. variability (across alts.):				
Mean std. dev.	1.12	0.11	0.82	1.31
Disp. of std. dev.	0.33	0.08	0.07	0.58
(2) Across-alternative attrib. variability (Ad hoc)				
Std. dev. of montly costs	0.81	0.62	0.06	2.87
" risk difference	1.1	0.34	0.47	1.53
" latency	0.96	0.46	0.09	2.42
" years sick	0.82	0.66	0	4.28
" lost life years	0.85	0.6	0	3.19
(3) Across-alternative attrib. variability (Structural)				
Std. dev. of linear net income term	0.85	0.59	5e-3	2.95
" quadratic income term	0.69	0.69	0.01	6.92
" $\Delta\Pi_i^{AS}\log(pdvi_i^A+1)$	0.89	0.59	0	3.16
" $\Delta\Pi_i^{AS}\log(pdvr_i^A+1)$	0.52	0.87	0	6.18
" $\Delta\Pi_i^{AS}\log(pdvl_i^A+1)$	0.89	0.61	0	3.19
Observable proxies of sociodemographics				
Income (in $1000)	50.27	33.5	5	150
Age	50.7	15.2	25	93
1(Female)	0.52	-	0	1
1(Divorced)	0.11	-	0	1
1(Black)	0.09	-	0	1
1(Other ethnicity)	0.04	-	0	1
1(Hispanic)	0.06	-	0	1
Household size	2.57	1.26	1	8
# of kids	0.52	0.95	0	5
1(Dual income household)	0.65	-	0	1
1(Single parent)	0.02	-	0	1
Observable proxies measures of cognitive capacity				
1(Less than high school)	0.11	-	0	1
1(High school degree)	0.34	-	0	1
Avg. duration on other choice occasions	45.97	26.4	0	202
1(Valid duration)	0.99	-	0	1
Attention behavior controls				
1(All status quo)	0.15	-	0	1
1(No change in difficulty rating)	0.21	-	0	1
Survey-specific health characteristics				
Illness experience count (0-13):	9.08	3.78	0	13
Avg. subj. risk of future experience (0-4):	-0.24	0.86	-2	2
Subjective controllability or risks (0-4):	-0.3	1.02	-2	2
1(Missing health)	0.09	-	0	1

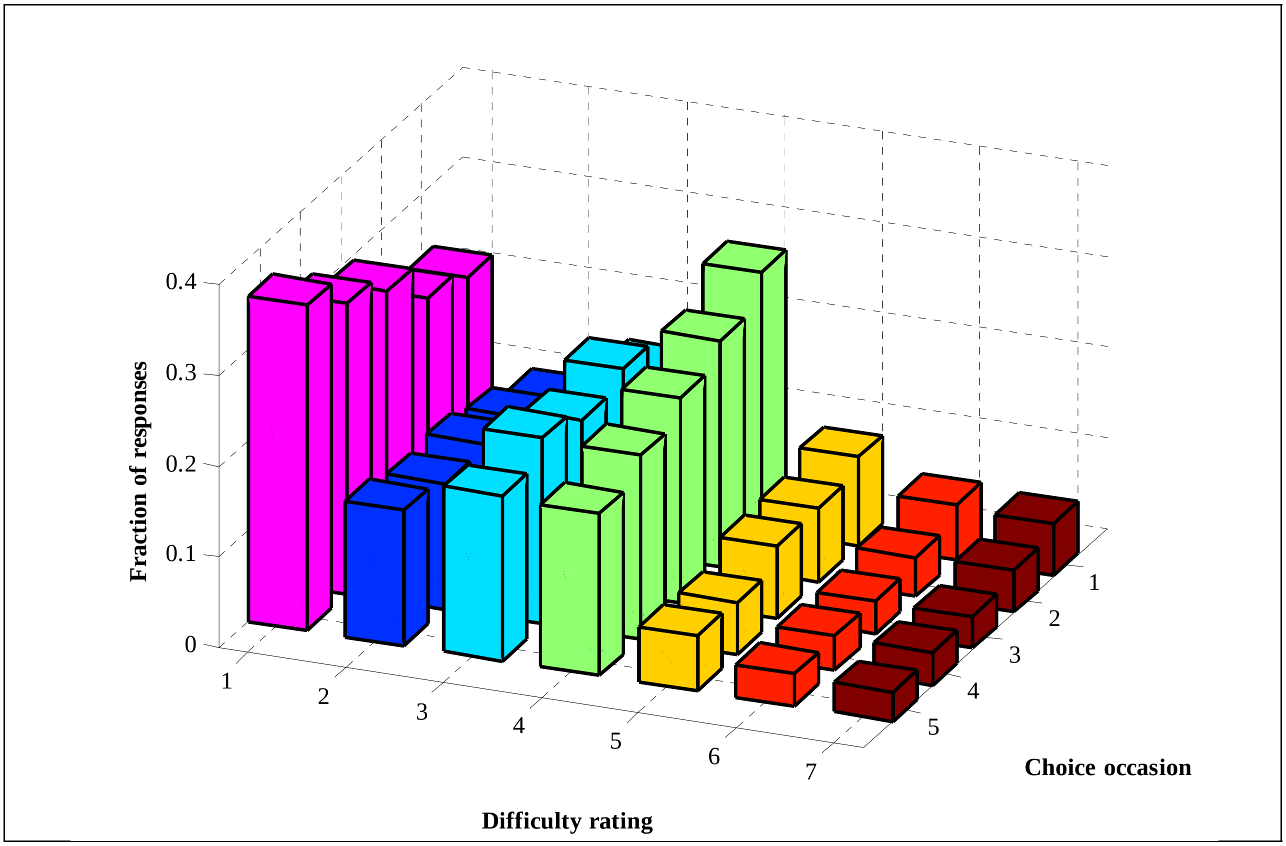


Figure 4 – Subjective Choice Difficulty Response Frequencies by Choice Occasion

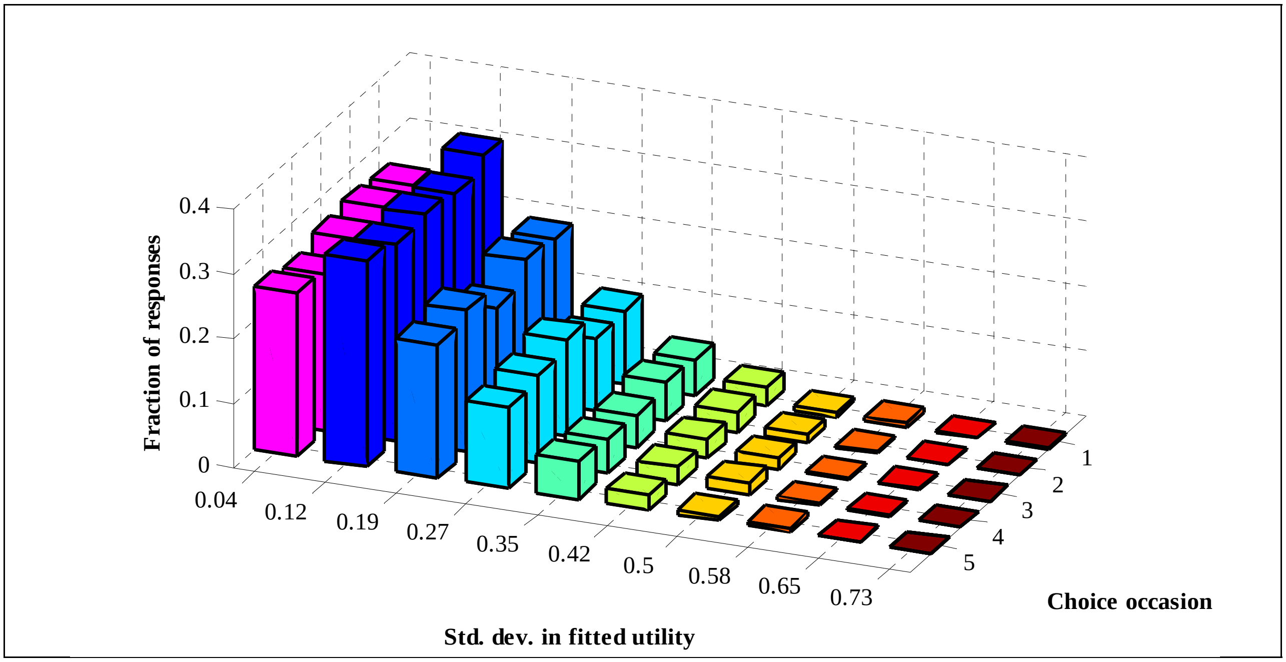


Figure 5 - Pattern of *Std. dev. of fitted U* by Choice Occasion

the standard deviation in attribute levels (e.g. $Mean\ SD_{i\bullet k}$ and $Disp.\ SD_{i\bullet k}$), and the standard deviation across alternatives of each program attribute (e.g., S*td. dev. of latency*). Furthermore, we include two sets of measures of across-alternative attribute-level correlations, corresponding to two different representations for how utility depends on program attributes—one ad hoc, and one more structural.

The “ad hoc” program attributes consist of the unprocessed attribute levels taken directly from the survey’s choice sets. These attributes include the monthly program cost, the size of the risk reduction, the latency of the illness, its duration, and the lost-life years. In our ad hoc specification, these raw attributes enter as linear and additively separable determinants of indirect utility. For our more-“structural” model, the program attributes are first processed to permit a structural random utility model within a formal discounted expected utility framework. Appendix A provides specific details on the construction of the typical DeShazo/Fermo-type objective complexity measures.

We employ a number of observable sociodemographic variables to explain differences in the subjective difficulty of choice tasks. These include age, gender (*Female*), marital status (*Single*, *Divorced*), race (*Black, Hispanic,* and *Other ethnicity,* relative to the omitted category *White*), number of household members, number of children in the household, an indicator for single parenthood, income, and an indicator for a dual-income household.

We also have a variety of health history variables for each respondent, as well as other subjectively reported variables. We make use of individuals’ subjective reports about their prior experiences with each class of illness, their subjective risk of suffering a future episode of each class of illness, and their perceptions about the subjective controllability of each type of illness. To accommodate occasional instances of missing health data, we construct an indicator variable, 1(*Missing health*), which has a value of one (zero otherwise) to identify these individuals.[13] To contain the dimensionality of the parameter space, we use only the individual's mean ratings across the list of illnesses for both the subjective risk measure and the subjective controllability measure.[14] To quantify each individual's personal experience with the major illnesses addressed in the survey, we introduce an individual-level variable which provides a simple count of the number of major illnesses the respondent indicates he or she has already experienced.

Our proxies for cognitive capacity include indicators for the highest level of education attained by the individual (i.e. 1(*Less than h.s.*) and 1(*High school*) (i.e. earned diploma), relative to the omitted category, at least *Some College*). We also include a measure of average time-on-task. To minimize endogeneity, this average is calculated for the respondent's other choice tasks, not including the choice in question (*Avg. duration on other choice occasions*).[15] This is consequently a choice-set-specific variable, since the nature of the "other" choice occasions will vary from choice to choice for an individual. Finally, there are some choice sets in the data for which the response time of individuals is exceedingly long. In many cases, this is probably because the individual took a break in the process of completing the survey. We handle these occurrences with an indicator variable, 1(*Valid duration*), which takes on a value of zero for exceedingly long response times and one otherwise. We interact this variable with the data on choice durations and use only duration data judged most likely to be valid in calculating the average time-on-task variable.

[13] Information on some or all of the health variables is missing for 166 individuals (or 812 choice sets) because these individuals chose not to respond to some health questions in Module 1 of the survey.
[14] Models where we use the disaggregated subjective responses of each illness type, instead of the mean value for these variables, provide qualitatively similar results.
[15] In empirical results not reported, we find that the use of current choice set response duration as a time-on-task measure to be positive and highly significantly correlated with choice difficulty, suggesting a strong endogeneity between the two measures. Also, we recognize that *Avg. duration on other choice occasions* may not completely mitigate the concerns of endogeneity bias because of the potential for joint dependence across choice occasions.

Several features of the raw distribution of the *How difficult* variable, as displayed in Figure 4, merit discussion. This figure highlights some of our concerns about the stability of preferences in a multiple choice-occasion stated-preference environment and suggests the likely need for additional control variables. First, the distribution of subjective ratings appears to be approximately normal, except for a mass of observations associated with the easiest difficulty rating (*How difficult=1*) on each of the five choice occasions. We believe that this heaping at "1" suggests that some proportion of our respondents may devote little attention to the question about their subjective difficulty rating. While they may engage sufficiently with the substantive program choice question, they may also recognize that the difficulty rating question is not as important and automatically choose the left-most option so that they can proceed more quickly through the survey.[16]

Another prominent feature of the distribution in Figure 4 is that respondents, on average, tend to rate their choices as being easier on each subsequent choice occasion. We thus introduce indicators for choice occasions two through five and treat the first choice for each respondent as the baseline throughout our empirical analysis.

To further explore the question of inattentive behavior, Table 2 displays the distribution of subjective difficulty ratings across all 26,451 choice occasions. For each difficulty rating, the table also displays the number and proportion of individuals who use the *identical* difficulty rating for all of their choices. A disproportionate share of responses for the "easy" rating—roughly forty percent—are from respondents who maintain the same rating across all five choice occasions. Of course, these individuals cannot express increasing ease of choices because they began at the "easy" end of the bounded scale. The other sixty percent of responses, however,

[16] Here, it would have been helpful to randomize the left-to-right order of the difficulty rating, sometimes putting "easy" on the left, and sometimes putting "very difficult" on the left, to check for primacy effects.

Table 2 – Invariant Difficulty Ratings

Variable: *How difficult* (rating)	Total # of responses	# with no change in rating	% with no change
1	8310	3414	41.08%
2	3597	342	9.51%
3	4800	480	10.00%
4	5793	963	16.62%
5	1845	93	5.04%
6	1050	90	8.57%
7	1056	258	24.43%
Total	26451	5640	21.32%

come from respondents who alter their difficulty rating at least once across the sequence of choice occasions.

If inattentive behavior is a consequence of choice difficulty, then the estimated marginal effects of each of the determinants of choice difficulty (such as objective choice set complexity) may suffer from a type of "attention" bias. We control for possible respondent inattention with an indicator variable, *All status quo*, for those individuals who always choose "Neither Program" for their conjoint choices. We also use an indicator, *No change in difficulty rating*, for those respondents who report the identical difficulty rating for all choice occasions. However, Malhotra (2009) finds evidence that inattention is more likely in the case of simple tasks ("survey satisficing"), and that people are "more motivated to persist in completing tasks [which are] intricate, challenging, and enriching." Thus we cannot automatically assume that increased difficulty leads to less attention to a choice problem.

4. Empirical Models

Before we can explore our models to explain subjective choice difficulty, we need to estimate some approximate utility parameters from a preliminary conditional logit choice model. These are needed so that we can build the "fitted" utilities required to construct the utility-space choice complexity measures— *Entropy* and *Std. dev. in fitted U*. These key measures, along with our other potential determinants of difficulty, are then used in the main model to explain respondents' subjective difficulty ratings for each choice task.[17]

[17] Work in progress includes the rather daunting task of developing a joint model that simultaneously uses respondents' reported choice difficulty ratings to shift the estimated preference parameters (and/or the scale factor) in our choice models. Here, we concentrate specifically on the determinants of perceived choice difficulty.

4.2 Preliminary estimation of utility function parameters

We consider two different specifications for the preliminary utility model from which we construct our utility-space measures of complexity. The ad hoc model has been outlined above. It merely uses the main raw attributes of each choice scenario to build a linear and additively separable utility "index" for construction of the utility-space measures. Our structural model, which borrows heavily from previous research with this same survey sample, allows us to construct utility-space measures based on a discounted expected utility specification. In Appendix B, we review in some detail the construction of the structural program attributes.

4.2 Models for Subjective Choice Difficulty

For individual *i* on choice occasion *t*, we model the subjective choice difficulty (*How difficult,* d_{it}) using a seven-interval ordered probit specification. Our goal is to assess the extent to which respondents' subjective choice difficulties are affected by objective measures of choice set complexity, by observable individual characteristics, and by apparent cognitive constraints,. We allow the latent continuous subjective choice difficulty, d_{it}^* , to be a linear-in-parameters function of several types of determinants:

$$d_{it}^* = \delta' w_{it} + \beta' x_i + \varepsilon_{it} , \tag{1}$$

where $i = 1,\ldots, N$ respondents and $t = 1,\ldots,T$ choice occasions per respondent. The vector w_{it} contains several objective measures of choice set complexity. The vector x_i captures a number of observable sociodemographic characteristics and proxies for cognitive capacity, which are assumed to be invariant over choice occasions for the same respondent. The error term, ε_{it} , is both individual- and choice-occasion specific. To identify the parameter vectors δ and β for the observable determinants of subjective choice difficulty, we assume that ε_{it} is distributed $N(0,\sigma_{\varepsilon}^{2})$ and that ε_{it} is uncorrelated with w_{it} and x_i for all individuals and all choice occasions.

The relationship between the observable ordered categorical response (represented by the individual's subjective difficulty rating, d_{it}) and the continuous latent difficulty variable is:

$$d_{it} = j \text{ if } \mu_{j-1} < d_{it}^{*} < \mu_{j} \tag{2}$$

where $j = 1,\ldots,7$, $\mu_0 = -\infty$, $\mu_7 = \infty$ and the other cut points $\mu_1,\ldots,\mu_6$ are estimated thresholds from the ordered probit regression analysis. Under the assumption that the error term is normally distributed, the probability of observing response $d_{it} = j$, conditional on w_{it} and x_i , is:

$$P_{itj} = P(d_{it} = j \mid w_{it}, x_i) = \Phi(\mu_j - \delta' w_{it} - \beta' x_i) - \Phi(\mu_{j-1} - \delta' w_{it} - \beta' x_i)\,, \tag{3}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function.[18]

As usual for discrete outcome models, neither the location nor the scale of the latent difficulty variable is known. Parameter identification requires the normalization to zero of either the intercept term in the latent regression specification, or one of the cut points (i.e. the thresholds between ratings). We make the assumption that the average intercept across individuals and choice sets is zero. Furthermore, to accommodate the arbitrariness of scale, the estimated coefficients, the cut points, and the error term are all normalized on the error standard deviation, e.g. $\delta = (\delta^{*}/\sigma_{\varepsilon})$, and the error variance for the normalized model is thus equal to one.

For completeness, we also consider a fixed-effects specification. In this context, we will assume that subjective choice difficulty, as measured by the difficulty ratings, is a cardinal

[18] Note that the ordered probit model is readily able to accommodate the heaping of responses that we observe at the easiest difficulty rating.

variable. In a least-squares framework, we introduce non-zero individual fixed effects, α_i , to parameterize the model for unobserved heterogeneity that is specific to each individual.

$$d_{it}^{*} = \delta' w_{it} + \alpha_i + \varepsilon_{it} \tag{4}$$

The fixed effects, α_i , will subsume all factors which are invariant over choice occasions for the same individual, so the β coefficients on the x_i variables in equation (1) cannot be separately identified. While the cardinality assumption is a compromise, this is one way to determine whether any unobserved heterogeneity might be biasing the estimated slopes on the remaining w_{it} variables (specifically, those variables which are not randomly assigned).

5. Results

Table 3 shows results for our two different specifications of the preliminary conditional logit models. These models produce estimates of the preference parameters for each attribute in a random-utility framework. Model 1 is our ad hoc specification of the latent utility index and includes only the raw attributes. In contrast, the structural model uses the constructed attributes described in Appendix B. This specification is reported as Model 2. Since the maximized log-likelihood for ad hoc Model 1 is larger than that for structural Model 2, we focus on choice set complexity measures in utility space and attribute space for the ad hoc model in our main illustration. However, we discuss a sensitivity analysis across the two possible models later in the paper.

Table 4 begins our estimation results for progression of variations on the ordered probit specifications in equation (2). Models 1 and 2, our most parsimonious specifications, provide striking evidence of the possible effects of choice complexity in utility space on subjective choice difficulty. Model 1 treats the *Std. dev. of fitted U* as the only determinant of respondents'

Table 3 – Simple preliminary conditional logit models[a]

COEFFICIENT	Model 1	Model 2
Ad hoc attributes:		
Annualized costs	-0.007***	-
	(-9.29)	
Risk difference	-50.920***	-
	(-4.40)	
Latency	0.002	-
	(1.30)	
Years sick	0.009***	-
	(3.92)	
Unexpected lost life years	0.012***	-
	(7.27)	
Structural attributes:		
Linear net income term	-	5.355***
		(9.19)
Quadratic net income term	-	-2.193***
		(-4.68)
$\Delta\Pi_i^{AS} \log(pdvi_i^A + 1)$	-	-24.793***
		(-4.23)
$\Delta\Pi_i^{AS} \log(pdvr_i^A + 1)$	-	-22.166**
		(-2.37)
$\Delta\Pi_i^{AS} \log(pdvl_i^A + 1)$	-	-30.717***
		(-6.02)
Observations	22485	22485
LogL	-11662.73	-11687.13

[a] For three-way choices between Program A, Program B, and Neither Program (N). z statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 4 - Utility-Space Determinants of Choice Difficulty[a]

COEFFICIENT	Model 1 Ordered Probit	Model 2 Ordered Probit	Model 3 Ordered Probit
Measures of choice set complexity:			
(a) In utility space:			
Std. dev. of fitted U	-0.353**	-	0.355
	(-2.40)		(0.61)
Entropy	-	3.094***	5.831
		(2.64)	(1.26)
Number of choice sets	7392[b]	7392	7392
Number of respondents	1694	1694	1694
Maximized LogL	-12851.805	-12851.200	-12851.014

[a] The intercept is normalized to zero for ordered probit models. Incidental threshold parameter estimates are not reported.
[b] Sample size has been reduced somewhat to match smaller sample with complete data for all of the variables in the models in Table 6.

subjective choice difficulty. Model 2 employs the calculated *Entropy* of the choice set as the only determinant. The results show that, independently, each utility-space measure is a strongly statistically significant determinant of the difficulty ratings selected by respondents (p<1%). The maximized value of the log-likelihood is also essentially identical across the two specifications.[19]

When we include *both* the *Std. dev. of fitted U* and *Entropy*, as in Model 3, their individual effects are insignificant (p>10%). To illustrate why this occurs, Figure 6 reveals that there exists a very close, although somewhat non-linear, relationship between these two utility-space measures. Thus, we are unable to distinguish between their separate effects when both variables are included in one model, and specifications like Model 3 are unhelpful.[20]

Table 5 preserves the utility-space *Entropy* variable from Table 4, but introduces two other attribute-space measures of objective choice set complexity as further explanatory variables for subjective choice difficulty. In section (b)(1) of this table, Model 2a reveals that the mean across alternatives of the within-alternative standard deviation of (standardized) attribute levels has a statistically significant negative effect on perceived difficulty.[21] A low mean value for these measures means that alternatives tend to have levels of attributes that are either all good, all bad, or all neutral, rather than mixes of attributes with some good and some bad, necessitating more tradeoffs during the decision process. We probably expect choices to be easier when the *Mean std. dev.* is small, and harder when more tradeoffs must be considered, which appears not to be the case. Model 2b, on the other hand, suggests that *Disp. of std. dev.*,

[19] In results not shown, we extend the linear specification of these variables to a quadratic form. The linear component for *Std. dev. of fitted U* is unchanged in regards to magnitude, sign, and significance, but the additional quadratic term is insignificant. Both linear and quadratic terms are insignificant when the specification of subjective difficulty is quadratic in *Entropy*.

[20] We perform nested likelihood-ratio tests of the restrictions present in Models 1 and 2 against the unrestricted model of Model 3. Confirming the Wald-type test embodied in the individual asymptotic t-test statistics on each parameter, these tests fail to reject the hypothesis (p>10%) of a zero incremental contribution for either variable when the other is already present in the model.

[21] See Appendix A.2 for a detailed exposition of how this objective complexity variable is calculated.

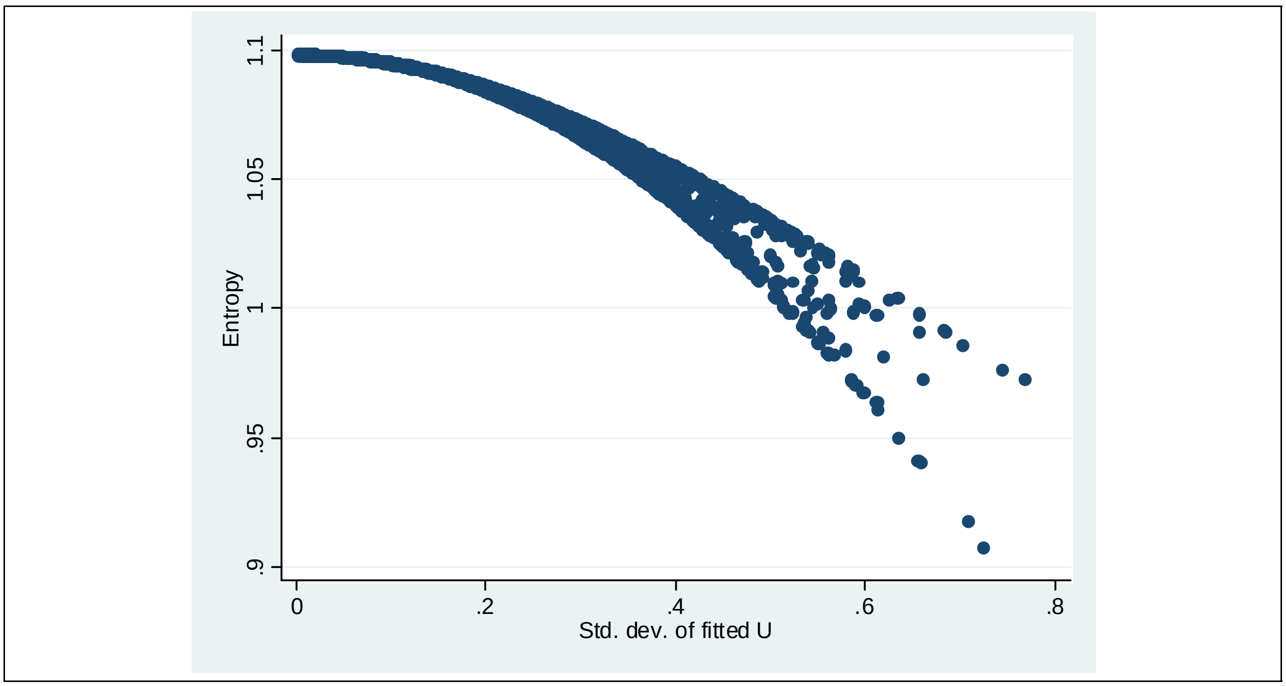


Figure 6 – Relationship between *Entropy* and *Std. dev. of fitted U*

Table 5 - Utility-Space and Attribute-Space Determinants of Choice Difficulty[a]

COEFFICIENT	Model 2a Ordered Probit	Model 2b Ordered Probit	Model 2c Ordered Probit	Model 2d Ordered Probit
Measures of choice set complexity:				
(a) In utility space:				
Entropy	3.202***	3.139***	3.202***	2.948**
	(2.73)	(2.68)	(2.73)	(2.50)
(b) In attribute space:				
(1) Distrib. of within-alt. attrib. variability				
Mean std. dev.	-0.304***	-	-0.304*	5.068**
	(-2.69)		(-1.81)	(1.97)
(Mean std. dev.)^2	-	-	-	-2.401**
				(-2.09)
Disp. of std. dev.	-	0.322**	-0.002	-
		(1.99)	(-0.01)	
Number of choice sets	7392[b]	7392	7392	7392
Number of respondents	1694	1694	1694	1694
Maximized LogL	-12847.582	-12849.212	-12847.582	-12845.396

[a] The intercept is normalized to zero for ordered probit models. Incidental threshold parameter estimates are not reported.
[b] Sample size has been reduced somewhat to match smaller sample with complete data for all of the variables in the models in Table 6.

the dispersion, across alternatives, of these same standard deviations (if added on its own) has a positive and statistically significant effect on perceived difficulty. In this case, some alternatives would have all good, all bad, or all neutral attribute levels, while others would have mixes of good and bad attribute levels. Choices appear to be judged more difficult when this is the case.

In Model 2c in Table 5, adding both the mean and dispersion of these standard deviations to this same model leaves only the mean term statistically significantly different from zero, so some of the information in these two measures appears to be duplicative. Model 2d, however, reveals that perceived difficulty is not linear in the *Mean std. dev.* characteristic of a choice set. This variable enters quadratically with a negative coefficient on the squared term. As the *Mean std. dev.* of within-alternative attribute levels increases from its minimum of 0.82 to its maximum of 1.31, perceived choice difficulty first increases, is maximized at a value of 1.06 for this variable, then decreases. The negative effect thus dominates if only a linear term is used, as revealed in Model 2a.

However, the models in Table 5 neglect other factors which may help to explain the variation in subjective difficulty ratings across individuals and across choices. If these other determinants are correlated with the *Entropy* variable (or with the *Std. dev. of fitted U* variable), then its coefficient may be biased. We check for this possibility, using just the *Entropy* variable as a utility-space measure, in the additional models presented in Table 6.

Section (b)(1) of Table 6 includes controls for both the *Mean std. dev.* and the square of *Mean std. dev.* as suggested by the results in Table 5. Section (b)(2) of Table 6 then summarizes the effects on perceived difficulty of standard deviations in attribute levels across alternatives on an attribute-by-attribute basis. Attribute levels are randomly assigned, except for occasional implausibility exclusions, so we expect no multicollinearity in these standard deviation

Table 6 –Additional Determinants of Subjective Choice Difficulty [a]

COEFFICIENT	Model 4 Ordered Probit	Model 5 Ordered Probit	Model 6 LS Fixed Effects
Measures of choice set complexity:			
(a) In utility space:			
Entropy	2.844**	4.077***	4.542**
	(1.99)	(2.58)	(2.42)
(b) In attribute space:			
(1) Distribution of within-alt. attrib. variability			
Mean std. dev.	4.842*	4.873*	6.822**
	(1.87)	(1.86)	(2.24)
(Mean of std. dev.)^2	-2.294**	-2.247*	-3.111**
	(-1.99)	(-1.93)	(-2.29)
(2) Across-alt. attrib. variability (Ad hoc model)			
Std. dev. of annualized costs	0.028	0.092***	0.073*
	(1.00)	(2.98)	(1.89)
" risk difference	-0.051	-0.063	0.053
	(-1.35)	(-1.64)	(1.19)
" latency	0.034	-0.080	-0.061
	(1.07)	(-1.59)	(-1.04)
" years sick	0.071***	0.032	0.045*
	(3.64)	(1.42)	(1.72)
" lost life years	-0.015	-0.051	-0.029
	(-0.57)	(-1.62)	(-0.78)
(3) Choice occasion indicators			
2nd choice occasion	-	-0.148***	-0.332***
		(-3.89)	(-8.30)
3rd "	-	-0.283***	-0.342***
		(-7.35)	(-8.40)
4th "	-	-0.350***	-0.372***
		(-9.00)	(-8.98)
5th "	-	-0.435***	-0.479***
		(-11.13)	(-11.53)
Obj. measures of cognitive capacity:			
1(Less than high school)	-	0.128***	n/a
		(2.89)	
1(High school degree)	-	0.060**	n/a
		(2.07)	
Avg. duration on other choice occasions	-	0.002***	-0.033***
		(4.39)	(-16.78)
1(Valid duration)	-	-0.504***	-1.481***
		(-3.62)	(-8.14)
Obj. measures of sociodemographics:			
Income (in $1000)	-	0.001**	n/a
		(2.01)	
Age	-	-0.018***	n/a
		(-2.68)	
Age^2	-	0.009	n/a

		(1.56)	
1(Female)	-	0.046*	n/a
		(1.85)	
1(Divorced)	-	-0.024	n/a
		(-0.56)	
1(Black)	-	-0.185***	n/a
		(-4.05)	
1(Other ethnicity)	-	-0.004	n/a
		(-0.06)	
1(Hispanic)	-	0.093*	n/a
		(1.74)	
Household size	-	0.023	n/a
		(1.17)	
# of children	-	-0.055**	n/a
		(-2.26)	
1(Dual income household)	-	-0.050*	n/a
		(-1.65)	
1(Single parent)	-	-0.048	n/a
		(-0.46)	
Survey-specific respondent characteristics:			
Illness experience count (0-13):	-	-0.010***	n/a
		(-2.83)	
Avg. subj. risk of future experience (0-4):	-	0.091***	n/a
		(5.65)	
Subjective controllability of risks (0-4):	-	0.046***	n/a
		(3.44)	
1(Missing health)	-	0.024	n/a
		(0.52)	
Respondent attention behavior:			
1(All status quo)	-	-0.068*	n/a
		(-1.72)	
1(No change in difficulty rating)	-	-0.632***	n/a
		(-18.70)	
Constant[b]	-	-	-2.438
			(-0.96)
Number of choices	7392	7392	7392
Number of respondents	1694	1694	1694
Maximized LogL	-12463.901	-10115.086	-10116.586

[a] All models are pooled ordered-probit except for Model 6, which uses a linear fixed-effects estimator and clusters on errors at the respondent level. Robust t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
[b] The intercept is normalized to zero for ordered probit models. Incidental threshold parameter estimates are not reported.

measures. Only the standard deviation across alternatives in the number of sick-years appears to have a positive and significant effect on subjective choice difficulty. These results conflict with our basic intuition that greater dissimilarity in an attribute should make alternatives easier, rather than harder, to compare. All other individual attribute-space standard deviation measures in the ad hoc specification bear coefficients which are statistically insignificant. These results, however, are for the case where we already control for the factors listed in Section (b)(1) of the table.

We note that the estimated effect of the *Entropy* variable changes only slightly between Models 1 and 4 despite the addition of the seven additional attribute-space measures. Overall, these results suggest that respondents' perceptions of choice difficulty are sensitive to the proximity of the alternatives in utility space as well as to the mix of objective attributes within a choice set (in ways that may be independent of preferences over these attributes).

Due to the essentially randomized design of the illness profiles, our measures of objective choice set complexity are orthogonal to all of the sociodemographic variables. Thus our models can in principle be estimated without controls for sociodemographic characteristics, without concerns about omitted variables bias in the coefficients of any of the purely attribute-space variables. However, we extend the specification in Model 5—to include indicators for choice occasions 2 through 5 as well as a range of sociodemographic variables—to see whether these variables further increase our ability to predict subjective difficulty ratings. Model 5 also includes our observable proxies for cognitive capacity, health history and subjective health variables, and some other controls that may capture inattention to the choice task.

In Model 5, we find that the coefficient on the *Entropy* becomes about one-third larger when we control for choice occasions and a wide range of respondent-specific characteristics. In section (b)(1) of the table, the coefficients on the quadractic form of *Mean std. dev.*change only slightly and maintain their significance. Section (b)(2) of the table reveals that respondents seem to view choices as *more* difficult if costs are more different across alternatives but with the standard deviation in years sick now becomes statistically insignificant. This asymmetry between the cost variable and the other illness profile attributes (such as years sick) may not be surprising, however.

Model 5 also allows us to identify the effects of important observable proxies for cognitive capacity and some observable sociodemographic characteristics of individuals. Among the cognitive capacity measures, we find a very clear gradient in the effect of education on subjective difficulty ratings. There is no significant difference in subjective difficulty between those individuals who completed college and those with only some college experience so we combine these as the omitted category. However, there exists a significant increase in rated choice difficulty for respondents who have only a high school degree compared to the baseline individuals who have at least some college education. This effect is even larger when the comparison involves individuals with less than a high school education.

Our proxy for other aspects of cognitive capacity (or constraints on its utilization), *Avg. duration on other choice occasions,* indicates that the net effects of these unobserved determinants, collectively, have a positive effect on rated choice difficulty. People who spent more time on other choice tasks tend to rate the current choice as more difficult.

Sociodemographic characteristics also influence subjective choice difficulty. Perceived choice difficulty seems to increase with income (which may actually measure the opportunity costs of time spent on these choice tasks) and to decline with the respondent's age (which may reflect either greater confidence about decision-making ability or more familiarity with health-risk related choices after controlling for educational attainment).[22] Perceived difficulty is also higher for females, lower for blacks but higher for Hispanics (relative to whites), and lower for households with more children. Membership in a dual-income household may correspond to lower perceived difficulty, although the mechanism for such an outcome is not clear.[23] Furthermore, subjective choice difficulty decreases with the number of illnesses with which a respondent has had prior personal experience. In contrast, larger values for the average subjective risk of future major illness (which tends to increase with age) and the average subjective controllability of illness correspond to greater subjective difficulty for choice tasks. These controls may offset some of the tendency of greater age to affect subjective choice difficulty.

Respondents are also more likely to rate a choice as being easier if they choose the status quo ("Neither Program") option for all the conjoint choices or report a constant subjective difficulty rating for all of the choice sets they considered. This last result supports with our conjecture that these particular respondents may have been relatively inattentive to the various choice tasks.

An obvious potential concern about Model 5 in Table 6, given that we have panel data in the form five choices for most individuals, is the possibility of bias from remaining unobserved heterogeneity. While the design of the offered attributes is randomized, so that we expect minimal correlation between these choice set design variables and any unobservable respondent

[22] We thought we might possibly identify an increase in choice difficulty for some of the oldest seniors, but perhaps selection bias among these oldest seniors means there are too few seniors in our sample who are old enough to be cognitively compromised to a statistically detectible extent. The point estimate of the coefficient on the square of age is positive, suggesting a U-shaped profile for perceived difficulty as a function of age, but the coefficient is not quite statistically significant at the 10% level.

[23] To the extent that choices are easy if the respondents simply checks "Neither Program" in every case, we have been careful to control for cases with this universal rejection of the offered programs. However, individuals who selected "Neither Program" in most cases, but not all, are not captured by this variable.

characteristics, it is possible that some of the observed respondent heterogeneity is correlated with unobserved heterogeneity. So we next explore a fixed-effects specification.

Model 6 in Table 6 involves a least-squares-based fixed-effects model to condition on the full set of choice-invariant characteristics, whether observed or unobserved, to better identify the effects of the choice-set varying factors.[24] Importantly, we find that the effect on subjective difficulty ratings of a change in the *Entropy* variable remains significant and negative. This particular measure of choice complexity in utility-space plays an important role in the conjoint choice responses of individuals—but not the *only* role.

We find in Section (b)(1) of Table 6 that the coefficients for the quadratic effect of the within-alternative attribute-space choice complexity measure, *Mean std. dev.*, retain their relative magnitudes. In section (b)(2) , the point estimates of the effects of the individual standard deviations of annualized costs and the illness profile attributes are relatively robust across Models 5 and 6 (with the exception of a sign change for the statistically insignificant effects of variability in the risk difference attribute).

Concerning the choice occasion indicators in Section (b)(3) of Table 6, Model 5 suggests that respondents rate choices as becoming successively easier, on average, with each additional choice. A test of equality of the estimated coefficients for the choice occasion indicator variables reveals significant differences across choice occasions. In Model 6, however, relative to the first choice, we fail to reject equality among the ratings differentials for choice tasks 2 through 4.

[24] We have also carried out the fixed-effects analysis with an unconditional fixed-effects ordered probit model that is available in the LIMDEP 9.0 software. The unconditional fixed-effects model is subject to the incidental parameters problem when the number of time periods (or choice sets) is small (see Greene (2008)). We did not find any qualitative changes in the relatives sizes of coefficient estimates when moving to this model, although we do find an upward shift (in absolute terms) in *all* the coefficients of the model, which is to be expected given our relatively small number of choice sets (i.e. five per respondent). Instead, we choose to report results for the simpler linear fixed-effects models under the assumption the individuals' apparent subjective difficulty ratings are approximately cardinal.

However, we can reject equality of the ratings differentials across choice tasks 2 through 5.[25] Thus, respondents (on average) report lower difficulty ratings across choice tasks but rated difficulty does not appear to decline in a linear fashion.

Finally, the sole striking difference in sign across Models 5 and 6 is for the coefficient on the average duration on other choice tasks for a respondent. In Model 5, choices were perceived as more difficult when a respondent tended to spend more time on other choices. This suggests that longer choice durations may reflect *lesser cognitive capacity*. Net of any unobserved heterogeneity, however, Model 6 suggests that longer choice durations correspond to judgments of *lesser* choice difficulty. This would be consistent with longer durations reflecting *lesser time constraints* on the decision-making process.

6. Discussion

Our results suggest that there are a number of important factors which influence the subjective difficulty of a choice task. We find that utility-distances between alternatives in the choice set clearly have a significant effect on the choice difficulty that respondents perceive. The *Entropy* variable and our alternative utility-space measure of these distances, *Std. dev. of fitted U*, appear to perform about equally well in capturing the choice difficulty that stems from the closeness of alternatives in utility-space.[26]

Our *Entropy* measure could, of course, be rendered more sensitive to differences in preferences across individuals if the preliminary conditional logit choice model involved greater

---

[25] Respondents were informed prior to their last choice occasion that the following choice would be their final one. This may explain the difference in the average difficulty rating of the last choice from the average ratings of the intervening choice occasions. Also, in an alternative specification, we incorporate choice occasion effects by using a linear index for the number of choice occasions. Under this alternative treatment, we include a linear and quadratic term for the index and find only the linear term (negatively) significant.

[26] In general, for data in which the number of alternatives is constant throughout the survey, any effects on choice difficulty from entropy can safely be attributed to the distances between alternatives in utility-space.

parameter heterogeneity. We have stayed with the simplest possible specifications in this model because some of our attribute-space variables (notably those under the heading of *Across-alt. attrib. variability*) are calculated on an attribute-by-attribute basis. However, there is no requirement that the list of attributes, or the functional form, in the preliminary choice model used to create the *Entropy* variable must be the same as the list of attributes introduced as regressors in the choice difficulty model. *Entropy* may, in fact, explain choice difficulty even better if the preliminary choice model is richer.[27]

Overall, our analysis of the effects of choice context suggest that the determinants of perceived choice difficulty likely extend well beyond just the simple proxies based on measures of objective choice complexity which have typically been explored in the existing literature. In addition to our alternative measures of proximity in utility space, subjective choice difficulty appears to vary systematically with a variety of dimensions of individual heterogeneity (e.g. income, age, ethnicity, and the number of children present in the household).

Explicit information about subjective choice difficulty could be incorporated into a richer (and much more complicated) joint empirical model. Importantly, our results suggest that the empirical estimation of demand or *WTP* may be affected by a rather wide variety of factors that have not typically been accounted for in the choice complexity literature. In this paper, we have used a very crude preliminary discrete choice model merely to produce the initial estimates of the utility parameters needed to build any measure of alternative similarity in utility space. In principle, this sub-model could be estimated simultaneously with another sub-model to explain subjective choice difficulty. Actual or fitted choice difficulty could be used simultaneously to shift the utility parameters and/or the error variances in the choice model. This would allow for a much broader analysis of the direct effects of choice difficulty on *WTP*. Having focused here on

[27] Models which assess this possibility are currently being explored.

the details of the sub-model for the subjective difficulty measure, however, we leave this more comprehensive analysis for subsequent work.[28]

It is also possible that the results based on the ad hoc specification for the choice model may not carry over to specifications. To assess this possibility, we use the structural attributes of the theoretical model that we outline in Appendix B, which is a simplification of the model employed in Cameron and DeShazo (2009). Table 7 reproduces the main coefficient estimates for the ad hoc specification (Model 6 in Table 6) along with the estimates for the corresponding structural specification (Model 7). Both models are estimated using the linear fixed-effects estimator as an approximation. The comparable attribute-space measures bear very similar coefficients, but the use of the standard deviations of the structural variables, instead of the ad hoc variables, causes the coefficient on the key *Entropy* variable to fall by half.

Another consequence of controlling for individual fixed effects in Model 6 is that the effect of *Avg. duration on other choice tasks* changes from significant and positive, to significant and negative. This generally implies that unobserved heterogeneity across individuals is likely correlated with average processing times for these types of choices. A binding time constraint may be one such omitted variable. We cannot control directly for how binding respondents' time constraints might have been. Longer observed response times might correspond to an opportunity for a more leisurely consideration of the alternatives, which might result in a perception of less difficult choices. Fischer, et al. (2000) note that while choice set complexity is likely to lead to longer response times, the observed pattern will be confounded if individuals

[28] We have research already in progress concerning this joint model. It is straightforward to specify such a model. However, because the estimated utility parameters show up in more than one place, convergence is difficult to achieve in a full information maximum likelihood context. We have had success with a model that alternates between (1) a logit model involving parameters and/or the error variance expressed as functions of the fitted values from the previous iteration of the difficulty model, and (2) an ordered logit subjective difficulty model conditional on fitted logit parameters from the previous iteration of the choice model. Iterating between the two conditional models permits convergence.

Table 7 – Comparison of Ad hoc and Structural Specifications [a]

COEFFICIENT	Model 6 LS Fixed Effects	Model 7 LS Fixed Effects
Measures of choice set complexity:		
(a) In utility space:		
Entropy	4.542**	2.729***
	(2.42)	(2.58)
(b) In attribute space:		
(1) Distribution of within-alt. attrib. variability		
Mean std. dev.	6.822**	7.476**
	(2.24)	(2.47)
(Mean std. dev.)^2	-3.111**	-3.403**
	(-2.29)	(-2.52)
(2) Across-alt. attrib. variability (Ad hoc model)		
Std. dev. of annualized costs	0.073*	-
	(1.89)	
" risk difference	0.053	-
	(1.19)	
" latency	-0.061	-
	(-1.04)	
" years sick	0.045*	-
	(1.72)	
" lost life years	-0.029	-
(3) Across-alt. attrib. variability. (Structural model)		
Std. dev. of linear net income term	-	0.055
		(0.99)
" quadratic net income term	-	-0.001
		(-0.02)
" $\Delta\Pi_i^{AS}\log(pdvi_i^A+1)$	-	0.072***
		(2.77)
" $\Delta\Pi_i^{AS}\log(pdvr_i^A+1)$	-	0.016
		(0.95)
" $\Delta\Pi_i^{AS}\log(pdvl_i^A+1)$	-	-0.009
		(-0.34)
Choice occasion indicators	Yes	Yes
Observable proxies for cognitive capacity	"	Yes[b]
Obs. measures of sociodemographics	"	n/a
Survey-specific health characteristics	"	n/a
Attention behavior controls	"	n/a
Number of choices	7392	7392
Number of respondents	1694	1694
Maximized LogL	-10115.086	-10116.586

[a] All models use a linear fixed-effects estimator and cluster on errors at the respondent level. Robust t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1. The full set of results is available from the authors.

[b] Valid duration indicator and average duration on *other* choice occasions only, to minimize endogeneity..

endogenously adopt decision strategies in response to the level of complexity in a fashion similar to those types of behaviors modeled in the effort-accuracy literature (e.g., Payne (1993)). If decision strategies are flexible, then response times could decrease for a given level of difficulty, leaving the general relationship between choice difficulty and response times ambiguous.

Every subject in our study was presented with only five choice occasions, which prevents us from effectively exploring, in any depth, some of the potentially confounding effects of evolving endogenous decision strategies on the relationship between response times and choice difficulty. However, we control crudely for these possible changes in the average individuals' response strategy with the choice occasion indicators in our empirical analysis. In general, this ambiguity suggests that future conjoint choice survey research may benefit if respondents were also asked about the extent to which they had to rush to make their choices.

We also hypothesize in our study that stated subjective choice difficulty potentially captures all of the different things that can conspire to make a particular stated preference choice situation "difficult" from the perspective of the individual respondent. Furthermore, subjective choice difficulty may differ across respondents even for identical choice tasks.[29] However, a potential concern with the use of subjective assessment of choice difficulty is that respondents may lack the experience necessary to properly locate the difficulty of the initial choice on any absolute scale, which may distort coefficient estimates for any of the factors of choice difficulty associated with the context and/or design of the survey.

[29] In a similar quest to our analysis, Luce, et al. (2003) extends the efforts by Fischer, et al. (2000) and allows half of the subjects the opportunity to put 90% confidence bounds on their initial ratings. The authors use these confidence bounds as subjective measures of the level of conflict that individuals consciously or unconsciously perceive. The response errors and confidence bounds are both shown to be affected by variation in attribute conflict, attribute extremity, and choice context.

Given that respondents have insufficient knowledge of the likely distribution of subjective choice difficulty for the *first* choice occasion, each respondent may select a difficulty rating for the first choice in a relatively arbitrary fashion. As respondents proceed through additional choice occasions, however, they begin to update their beliefs about the distribution of difficulty levels across choice occasions. In a survey containing a large enough number of choice occasions, the influence of the initial rating—affected by the respondents' prior belief about the distribution of possible choice difficulties and his or her guess about where the first choice may lie on the overall difficulty spectrum—might eventually disappear as respondents gain experience. One could omit the first choice tasks and their ratings from the analysis, treating them as part of a "burn-in" phase. However, our survey involves only five choice occasions per respondent and this limits our ability to fully address the possibility of these initial reference level effects. It may be possible to address this concern by normalizing on respondents' initial choice ratings, although the boundedness of the seven-point scale is a limitation.

7. Conclusions

Previous studies have not enjoyed the advantage of a directly elicited subjective difficulty rating for each one of a large set of stated choice tasks, with multiple choices per respondent, such as the atypical variable we exploit in this paper. As a result, existing studies have typically relied upon on only some of the many possible proxies for choice difficulty. "Choice difficulty" is often invoked as the latent factor which explains why some of these proxies have the systematic effects on marginal utilities or scale factors that they are observed to produce. However, it has only been possible to speculate that "choice difficulty" is the relevant missing link (i.e. unobserved mediating variable).

Without a specific choice difficulty variable, researchers who wish to control for choice difficulty need to be satisfied with controlling for it *indirectly* instead, using one or more raw or constructed quantities based upon observable variables. Each of these variables may be able to explain *some* of the variation in the missing choice difficulty variable, but none does it all. We have demonstrated this handicap by showing that several different classes of variables seem to be predictive of individuals' reported subjective choice difficulty ratings.

Our findings suggest that directly elicited subjective choice difficulty ratings may have the potential to serve as a sound univariate summary of these numerous determinants of choice difficulty. Thus, future stated and revealed-preference research may be able to circumvent the adoption of some of the more sophisticated empirical choice models (e.g., Louviere (2001), Swait and Adamowicz (2001b), Hensher (2004), Greene and Hensher (2007)) to account for factors in the choice environment that can bias parameter estimates. In particular, our relatively non-intrusive follow-up question about the difficulty of the preceding choice may reduce the need for a highly parameterized empirical choice model with many kinds of objective proxies for contextual determinants of choice difficulty. In future analyses, a direct measure of subjective choice difficulty may be a viable way to control for some or all of the potential effects on respondent behavior originating from the challenges of the choice environment, in general, or for different types of individuals.

References:

Adamowicz, W., DeShazo, J.R. (2006). "Frontiers in stated preferences methods: An introduction", *Environmental & Resource Economics* 34, 1-6.

Bernheim, B.D., Rangel, A. (2009). "Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics", *Quarterly Journal of Economics* 124, 51-104.

Beshears, J., Choi, J., Laibson, D., Madrian, B. (2008). *How are preferences revealed?* National Bureau of Economic Research, Inc.

Cameron, T.A., DeShazo, J.R. (2009). "Demand for health risk reductions", Department of Economics, University of Oregon Working Paper.

DeShazo, J.R., Fermo, G. (2002). "Designing choice sets for stated preference methods: The effects of complexity on choice consistency", *Journal of Environmental Economics and Management* 44, 123-143.

Evans, M.F., Flores, N.E., Boyle, K.J. (2003). "Multiple-bounded uncertainty choice data as probabilistic intentions", *Land Economics* 79, 549-560.

Fischer, G.W., Luce, M.F., Jia, J.M. (2000). "Attribute conflict and preference uncertainty: Effects on judgment time and error", *Management Science* 46, 88-103.

Greene, W.H. (2008). *Econometric analysis*. Upper Saddle River, N.J.: Prentice Hall.

Greene, W.H., Hensher, D.A. (2007). "Heteroscedastic control for random coefficients and error components in mixed logit", *Transportation Research Part E-Logistics and Transportation Review* 43, 610-623.

Haaijer, R., Kamakura, W., Wedel, M. (2000). "Response latencies in the analysis of conjoint choice experiments", *Journal of Marketing Research* 37, 376-382.

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

Hensher, D.A. (2006a). "How do respondents process stated choice experiments? Attribute consideration under varying information load", *Journal of Applied Econometrics* 21, 861-878.

Hensher, D.A. (2006b). "Revealing differences in willingness to pay due to the dimensionality of stated choice designs: An initial assessment", *Environmental & Resource Economics* 34, 7-44.

Hensher, D.A., Rose, J., Bertoia, T. (2007). "The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies", *Transportation Research Part E-Logistics and Transportation Review* 43, 73-89.

Hensher, D.A., Rose, J., Greene, W.H. (2005). "The implications on willingness to pay of respondents ignoring specific attributes", *Transportation* 32, 203-222.

Johnson, F.R. (2006). "Comment on "revealing differences in willingness to pay due to the dimensionality of stated choice designs: An initial assessment"", *Environmental & Resource Economics* 34, 45-50.

Lancsar, E., Louviere, J., Flynn, T. (2007). "Several methods to investigate relative attribute impact in stated preference experiments", *Social Science & Medicine* 64, 1738-1753.

Li, C.Z., Mattsson, L. (1995). "Discrete-choice under preference uncertainty - an improved structural model for contingent valuation", *Journal of Environmental Economics and Management* 28, 256-269.

Louviere, J., Street, D., Carson, R., Ainslie, A., Deshazo, J.R., Cameron, T., Hensher, D., Kohn, R., Marley, T. (2002). "Dissecting the random component of utility", *Marketing Letters* 13, 177-193.

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J.R., Fiebig, D., Greene, W., Hensher, D., Waldman, D. (2005). "Recent progress on endogeneity in choice modeling", *Marketing Letters* 16, 255-265.

Louviere, J.J. (2001). "What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon", *Journal of Consumer Research* 28, 506-511.

Luce, M.F., Jia, J.M., Fischer, G.W. (2003). "How much do you like it? Within-alternative conflict and subjective confidence in consumer judgments", *Journal of Consumer Research* 30, 464-472.

Malhotra, N. (2009). "Order effects in complex and simple tasks", *Public Opinion Quarterly* 73, 180-198.

Mas-Colell, A., Whinston, M.D., Green, J.R. (1995). *Microeconomic theory*. New York, NY: Oxford University Press.

Payne, J.W., Bettman, J. R., Johnson, E.J. (1993). *The adaptive decision maker*. New York: Cambridge Univ. Press.

Rose, J.M., Black, I.R. (2006). "Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments", *Marketing Letters* 17, 295-310.

Shannon, C.E. (1948). "A mathematical theory of communication", *Bell System Technical Journal* 27, 379-423 and 623-656.

Swait, J., Adamowicz, W. (2001a). "Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice", *Organizational Behavior and Human Decision Processes* 86, 141-167.

Swait, J., Adamowicz, W. (2001b). "The influence of task complexity on consumer choice: A latent class model of decision strategy switching", *Journal of Consumer Research* 28, 135-148.

Vossler, C.A., Ethier, R.G., Poe, G.L., Welsh, M.P. (2003). "Payment certainty in discrete choice contingent valuation responses: Results from a field validity test", *Southern Economic Journal* 69, 886-902.

Welsh, M.P., Poe, G.L. (1998). "Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach", *Journal of Environmental Economics and Management* 36, 170-185.