Cameron, Trudy Ann; DeShazo, J. R.

Article

Differential attention to attributes in utility-theoretic choice models

Journal of Choice Modelling

Provided in Cooperation with:
Journal of Choice Modelling

Suggested Citation: Cameron, Trudy Ann; DeShazo, J. R. (2010) : Differential attention to attributes in utility-theoretic choice models, Journal of Choice Modelling, ISSN 1755-5345, University of Leeds, Institute for Transport Studies, Leeds, Vol. 3, Iss. 3, pp. 73-115

Journal of Choice Modelling, 3(3), pp 73-115
www.jocm.org.uk

Differential Attention to Attributes in Utility-Theoretic Choice Models

Trudy Ann Cameron[1,*] J. R. DeShazo[2,†]

[1]Department of Economics, 435 PLC, 1285 University of Oregon,
Eugene, OR, 97403-1285 USA
[2]Department of Public Policy and Institute of the Environment, 3250 School of Public Affairs,
University of California, Los Angeles, CA 90095-1656 USA

Received 20 October 2008, revised version received 10 March 2010, accepted 1 November 2010

Abstract

We show in a theoretical model that the benefit from additional attention to the marginal attribute within a choice set depends upon the expected utility loss from making a suboptimal choice if it is ignored. Guided by this analysis, we then develop an empirical method to measure an individual's propensity to attend to attributes. As a proof of concept, we offer an empirical example of our method using a conjoint analysis of demand for programs to reduce health risks. Our results suggest that respondents differentially allocate attention across attributes as a function of the mix of attribute levels in a choice set. This behaviour can cause researchers who fail to model attention allocation to estimate incorrectly the marginal utilities derived from selected attributes. This illustrative example is a first attempt to implement an attention-corrected choice model with a sample of field data from a conjoint choice experiment.

Keywords: Attention to Attributes, Allocation of Attention, Conjoint Choice, Choice Set Design, Bounded Rationality, Choice Heuristics

* Corresponding author, T: +01-5413461242, F: + 01-5413461243, cameron@uoregon.edu
† T: +01-3105931198, F: + 01-3102060337, deshazo@ucla.edu

ISSN 1755-5345

1 Introduction

Simple empirical choice models assume that the investigator knows exactly what information the individual uses to make a given choice—i.e. that the individual fully attends to, and costlessly processes, all the information available within a choice scenario. Economists, and choice modellers more generally, now recognize that the constituent elements of attention, including cognition and time, are scarce resources which rational individuals should allocate optimally (Simon 1955; March 1978; Heiner 1983, 1985; de Palma et al. 1994; Conlisk 1996; Gabiax and Laibson 2000). The optimal allocation of attention across attributes of alternatives will depend upon both the expected marginal benefits and marginal costs of further information processing. As a consequence, prior to making a choice, the individual may rationally attend to some attributes of the alternatives more than others.

However, this process of optimally allocating attention over the array of information in a choice set may create a profound problem for discrete choice researchers. Suppose an individual does value the level of a particular attribute, but because of some resource constraint, overlooks differences in its levels across alternatives when making their choice. A random utility empirical model, based on perfect and costless information, will imply that the marginal utility associated with that attribute is zero. Similarly, incomplete attention to any particular attribute, could lead to biased estimates of the effect of variations in the level of that attribute on the individual's choice. In particular, the apparent marginal utility in this case would be an attenuated estimate of the true marginal utility under perfect and costless information.

Further complications may arise if the individual's allocation of attention differs systematically across particular types of attributes. For example, consider the case of estimating demand for goods in economic applications. In revealed preference contexts, some individuals might allocate a disproportionately greater level of attention to prices as opposed to other attributes. This might raise the estimated relative marginal utility of net income and thereby lower the estimated willingness to pay (WTP). In stated preference contexts, however, researchers are often worried that people will instead pay too little attention to an alternative's price, thus lowering the estimated relative marginal utility of net income and artificially inflating estimated WTP.

In this paper, we derive results based on pre-choice optimization behaviour which lead to guidelines for empirical specifications. We develop a theoretical model that motivates our methodological approach, which we then illustrate with an empirical example. We argue that the benefits from additional attention allocated to the evaluation of an incremental attribute stem from the expected value of the avoided lost utility associated with a suboptimal choice (made as a result of ignoring that incremental attribute). The expected magnitude of this utility loss depends upon two components. The first component is the "other-attribute utility dissimilarity." This component represents how close the alternatives are in utility space—given the other attributes evaluated thus far. The second component is the "own-attribute utility dissimilarity." This component captures how much of a difference it might make, to overall utility from each alternative, if the incremental attribute is taken into account.

Our theoretical model leads us to develop a practical implementation of its insights, so that empirical choice specifications can accommodate the individual's "propensity to attend" to each attribute in a choice scenario. Conceptually, this propensity to attend to an incremental attribute is identified based on individual-specific measures of other- and own-attribute utility dissimilarities as well as the cognitive costs of attribute evaluation. We introduce a multiplicative propensity-to-attend parameter for each attribute which can be

viewed equivalently as affecting either the apparent marginal utility associated with the marginal attribute, or the perceived difference in the level of this attribute across the two alternatives.

As a proof of concept, we offer an empirical example of our method using a conjoint analysis of demand for programs to reduce health risks. Our results suggest that the combination of other-attribute dissimilarity and own-attribute dissimilarity causes respondents to differentially allocate attention across attributes. More specifically, this process appears to result in a tendency for the researcher to overestimate the marginal utility derived from net income, overestimate the marginal disutility of sick-years and lost-life-years, but perhaps to underestimate the marginal disutility of recovered-years, on average.

Although there is certainly a great deal of room to expand the theoretical scope of our model, the empirical version of this simple attention-corrected model is important for several reasons. First, the attention-corrected model has the potential to identify and eliminate distortions in the estimated parameters of discrete choice models which are estimated using either revealed or stated preference data. Second, it provides a clear measurement framework to test emerging hypotheses from the behavioural literature about the determinants of an individual's allocation of attention. Third, when the prospective real consequences of a choice vary from context to context, we might also expect the individual's budgeted attention to vary as well. Our attention-corrected model might also be used to explain some types of observed differences across data generation methods, such as revealed preference (RP) versus stated preference (SP) information. Fourth, our model opens up the possibility of beginning to measure the effectiveness of agents (such as marketers, salespeople, politicians, etc.) who strategically seek to direct the individual's attention towards some attributes and away from others as they design choice sets. The extent, and implications, of such strategic behaviour on choice outcomes, and therefore upon individual welfare, cannot be assessed adequately without the benefit of a framework with features similar to those of the model presented here.

The scope of the theoretical model is modest; it emphasizes the role of expected marginal benefits in determining the allocation of attention, although we are careful to outline how costs could also be incorporated. An illustration that includes the role of cost information awaits richer data. The present paper also focuses only on the allocation of attention to the marginal *attribute*. We leave a similar analysis of optimal attention to the marginal *alternative* for a subsequent paper.[1]

1.1 Related literature

Economists have long recognized that individuals face various resource constraints as they acquire and deploy information in their decision-making processes (Simon 1955; March 1978; Heiner 1983, 1985; de Palma, et al. 1994; Conlisk 1996). Theoretical models that predict how individuals might optimally respond to such constraints have only recently begun to emerge (Gabiax and Laibson 2000) and be tested in experimental settings (Gabaix et al. 2006). DellaVigna (2009) reviews the state of the literature on psychology and economics and inventories a small literature wherein inattention is assumed to vary inversely with the salience of "opaque" information and directly with the number of competing stimuli, but this literature appears to emphasize the identification of types of opaque information to which decision makers are not fully attentive. While many advances

[1] The notion of individual-specific "consideration sets" of relevant alternatives has been addressed in Haab and Hicks (1999), Chakravarti and Janiszewski (2003), Paulssen and Bagozzi (2005) and Jedidi and Kohli (2005).

have recently been introduced that inform the design of attribute-based field studies of demand, no general methods exist for directly modelling the effects of the allocation of attention, within a choice task, on the estimated marginal utilities (or part-worths) for these attributes.[2]

In developing a directed cognition model, Gabaix and Laibson (2000; 2005) approximate the economic value of additional attention using two ideas stemming from the analysis of option values. First, the option value of continued consideration declines as one alternative gains a large edge over other available alternatives. In the context of the present paper, this corresponds to the case where alternatives come to be perceived as less similar in terms of the utility they generate (i.e. when one alternative clearly dominates the other(s) in terms of the current information set). Second, the option value also declines when continued consideration yields little new information. In the context of the present paper, this corresponds to the case where additional attributes are more similar across alternatives or as units of these additional attributes provide minimal marginal utility.

Overall, the experimental results in Gabaix et al. (2006) are consistent with two implications of their option value framework, which defines how many search operations the subject should pursue. Translating their discussion into the terminology of our paper, the value of [attribute] exploration [for a given alternative] decreases the larger the gap between the active [alternative] and the next best [alternative]. Second, the value of [attribute] exploration increases with the variability of the information that will be obtained (p. 1053). When extended to their yoked case, where an additional attribute is revealed simultaneously for all alternatives, these insights appear to be essentially equivalent to the ones we derive analytically in this paper, in the context of a conventional empirical random utility choice model.

The Gabaix et al. (2006) experimental design, couched exclusively in monetary amounts, eliminates the need to measure physical quantities of an attribute, to estimate the marginal utility associated with that attribute, or to infer the marginal WTP for units of each attribute by considering marginal rates of substitution between that particular attribute and money. Money-denominated attributes cannot differ in their salience across individuals, since utility is implicitly considered to map directly into the total number of cents paid. However, real choice situations in field settings are confounded by heterogeneity in marginal utilities across attributes, differences in attribute metrics, as well as differences in individuals' cognitive abilities and the opportunity costs of their time.[3]

Many advances have recently occurred with respect to the design of attribute-based field studies of demand but they fall short of the goal of modelling the allocation of attention.[4] Hensher and co-authors have initiated several intriguing explorations of attention within a standard multivariate discrete choice setting. Hensher et al. (2005) use a specific follow-up question about which attributes the respondent did not use in making his or her choices. Hensher et al. (2007) also uses the same follow-up question to identify nine distinct attribute processing rules in the same. Respondent adherence to these rules is modelled as stochastic. The authors then use a modified mixed logit model which conditions each

[2] However, Swait and Adamowicz (2001b) take to task the empirical choice modelling community for its persistence in assuming a "utility-maximizing, omniscient, indefatigable consumer."

[3] The choice experiments in Fischer et al. (2000b) use the same Mouselab software as in Gabaix et al. (2006), but the choice data from their study is not utilized in an econometric random utility model.

[4] Psychologists have explored how various within-choice-set conditions may increase the cognitive costs of attribute evaluation and comparison (Bettman et al. 1993, 1998; Dellaert et al. 1999; Fisher et al. 2000a,b; Luce et al. 2003; Johnson 2008). Similarly, marketing scholars and others have explored the effects of task complexity on choice outcomes (Shugan 1980; Malhotra 1982; Mazzotta and Opaluch 1995).

parameter on whether a respondent included or excluded an attribute in their information processing strategy. In their conclusions, these authors acknowledge that there may be differences "between what people say they think and what they really think" (p. 216), and they question whether the "simply conscious statements" made by survey respondents represent an adequate measure of information processing. They emphasize that individuals' information processing strategies "should be built into the estimation of choice data from stated choice studies" (p. 214). This is precisely what we endeavour to accomplish in the present paper.

Employing a similar research method, Hensher et al. (2006a) find that that probability of a respondent considering more attributes decreases as the attributes used in their survey are drawn from distributions with narrower ranges. However, attribute ranges in this study appear to be varied simultaneously across all attributes and the dependent variable is available only at the level of the individual, not the choice set. In contrast, our models lead us to consider differences in the ranges of attributes within a single choice set as additional potential determinants of attention, and therefore of apparent marginal utilities and ultimately our estimates of WTP. Very relevant to our study is Hensher's finding that individuals' processing strategies depend on the *nature* of the attribute information in the choice set, not just the *quantity* of such information (i.e. the number of attributes).

Another related study is Puckett and Hensher (2008). This study seeks to integrate the attribute processing strategies (APSs) reported by respondents into the analysis of choices. The relevant APSs involve certain attributes being ignored, or aggregated to a different extent, by different respondents. This work builds on Hensher et al. (2006a) in that it considers the effects of APSs utilized by respondents for every alternative in every choice set, including across choice tasks faced by a given respondent. This approach can accommodate cases where attribute level mixes are outside of the acceptable choice bounds for the individual. The wording of their debriefing question for each choice was: "Is any of the information shown not relevant when you make your choice? If an attribute did not matter to your decision, please click on the label of the attribute below. If any particular attributes for a given alternative did not matter to your decision, please click on the specific attribute." Subjective all-or-nothing attention to different attributes is thus elicited directly from each respondent, rather than being inferred from choice behavior.

Finally, there is large literature that explores how the design of choice sets affects individuals' choice consistency and willingness to pay. Early versions of these include Mazzotta and Opaluch (1995) and DeShazo and Fermo (2002), leading up to the very ambitious "design-of-designs" studies by Hensher (2006b). Much of this work, however, is motivated by a concern with how the cognitive costs of information processing vary with choice set design.

Our focus in the present paper concerns an optimizing model for how the expected benefit from additional information drives the allocation of attention across attributes. Several of these other papers emphasize how, through deliberate manipulation of choice set design, the researcher can alter the estimated parameters. We contend that even if all of the survey instruments in a study employ the identical choice set design (in terms of numbers of attributes, alternatives, choice sets, attribute levels, and ranges) there can still be artefacts of the researcher's design decisions—with respect to the mix of attributes in any given choice set—that can unintentionally or intentionally affect the recovered utility parameter estimates. Furthermore, these effects can vary across individuals.

2 A Theoretical Model for Attention to Attributes

Suppose subjects in a stated preference (SP) choice experiment (or in revealed preference (RP) choice data) actually do care about the level of a particular attribute. However, for some reason, they fail to evaluate its levels across alternatives when making their choices. In this situation, a random utility empirical model, based on perfect and costless information, will imply that the marginal utility associated with that overlooked attribute is zero. Likewise, simply incomplete attention to any particular attribute, as opposed to zero attention, could be expected to result in a lesser-than-expected effect of variations in the level of that attribute on people's choices. The apparent marginal utility in this case would be an attenuated estimate of the "true" marginal utility under perfect and costless information.

If the subject's cognitive resources are limited, but this "inattention effect" is uniform across attributes, then perhaps all of the indirect utility parameters may be proportionally attenuated. This is observationally equivalent to the case where the "scale factor" in a discrete-choice model is smaller (i.e. the error variance is larger). If the propensity to attend to attributes is scaled down equally for all attributes (including net income) when an individual is paying less attention, we would expect to see no bias created in the implied point estimate of marginal WTP for that attribute. The ratio of the marginal utility associated with any attribute, relative to the marginal utility of income, is typically all that matters in simple models.[5]

However, it is possible that inattention to attributes *differs* across attributes. In particular, when decision resources are limited, there may remain a disproportionate level of attention devoted to an alternative's cost as opposed to other attributes. In this case, distortions in WTP could be expected. Relatively less attenuation in the estimated marginal utility of income will inflate the denominator relative to the numerator in the usual WTP calculation, so that the implied WTP for every alternative could be biased toward zero. In contrast, in the context of SP models, there is great concern that because of the hypothetical nature of the choices involved, the subject will fail to pay sufficient attention to the cost variable. The emphasis on providing a "cheap talk" script as part of the survey is designed to draw the respondent's attention specifically to the cost variable and its implications. Disproportionately greater attention to the implications of the cost variable may serve to amplify attention to this attribute relative to other attributes when the expected utility loss is otherwise rather low because of the less directly consequential nature of many SP choices. If other attributes of the offered alternatives are not similarly emphasized, the practice of offering only a cheap talk script—generally intended to increase the apparent marginal utility of income—can be expected to produce a downward bias in estimated willingness to pay (relative to a scenario that worked equally hard to draw respondents' scarce attention toward all attributes).[6]

In this paper, we derive some results based on optimization behaviour which lead to guidelines for empirical specifications. Our models concern how the individual's optimal

[5] Conlon et al. (2001) use response time as a proxy for consumer effort devoted to a choice, where effort is regressed on choice set characteristics and involvement measures. One choice set characteristic is the expected utility difference across alternatives, based on a preliminary multinomial logit choice model.

[6] An anonymous referee has suggested that when attention to attributes is not scaled back proportionally for all attributes, the effect on a conventional choice model might be analogous to the introduction of an attribute-specific error term. However, we concentrate upon a possible structural interpretation involving the marginal benefits of attention, rather than relegation of this phenomenon to the stochastic structure of the model.

amount of attention to a particular attribute might be determined by the nature of the decision context and how this context interacts with the preferences of that individual.

2.1 A Two-Alternative Case

Consider first a familiar binary choice model (with alternatives indexed by 0 and 1) where the underlying indirect utility function is linear and additively separable in net income (i.e. $\left(Y_i - T_i^j\right)$ = income minus the cost of option *j*) as well as several other attributes, X_{ki}, $k = 1,...,K$.

$$\begin{aligned} V_i^1 &= \beta_1\left(Y_i - T_i^1\right) + \sum_{k=2}^{K}\beta_k X_{ki}^1 + \varepsilon_i^1 \\ V_i^0 &= \beta_1\left(Y_i - T_i^0\right) + \sum_{k=2}^{K}\beta_k X_{ki}^0 + \varepsilon_i^0 \end{aligned} \tag{1}$$

The utility-difference expression driving the choice between alternatives 1 and 0 can thus be written as:

$$\begin{aligned} V_i^1 - V_i^0 &= \beta_1\left(T_i^0 - T_i^1\right) + \sum_{k=2}^{K}\beta_k\left(X_{ki}^1 - X_{ki}^0\right) + \left(\varepsilon_i^1 - \varepsilon_i^0\right) \\ &= -\beta_1 t_i + \sum_{k=2}^{K}\beta_k x_{ki} + \varepsilon_i \end{aligned} \tag{2}$$

where $\left(T_i^0 - T_i^1\right) = \left(X_{1i}^1 - X_{1i}^0\right) = x_{1i} = -t_i$ is often distinguished from the other attribute differences because of the role of its coefficient, β_1, in the calculation of *WTP*. In general, lower-case variable names will be used to denote differences in attribute levels between alternative 1 and alternative 0 (i.e. "net" levels of attributes in this two-alternative context).

In this simple linear specification, *WTP* is calculated by setting the utility difference to zero and solving for the level of t_i^* which creates this indifference between alternatives. The implied *WTP* and E[*WTP*] functions are

$$\begin{aligned} WTP_i = t_i^* &= \frac{\left(\sum_{k=2}^{K}\beta_k x_{ki}\right) + \varepsilon_i}{\beta_1} \\ E\left[WTP_i\right] &= E\left[\frac{\sum_{k=2}^{K}\beta_k x_{ki}}{\beta_1}\right] + E\left[\frac{\varepsilon_i}{\beta_1}\right] \end{aligned} \tag{3}$$

2.1.1 Marginal Benefit of Attention to an Additional Attribute

The marginal benefit to the individual of paying attention to an additional attribute can be equated to the avoided expected utility loss from making an incorrect choice as a result of ignoring X_k in the choice process. There are two ways that the subject can experience a loss from failing to consider X_k. First, she might choose alternative 1, when in fact (i.e. on the basis of the full set of attributes) her utility would actually be higher under alternative 0. Or, she might choose alternative 0 when her utility would actually be higher under alternative 1,

if all attributes were taken into account. Her expected utility loss from failing to consider the level of X_k will be:

$$\begin{aligned} E\left[U\ Loss\right] &= \Pr\left[1\ chosen \mid 0\ optimal\right]\left(V_i^0 - V_i^1\right) \\ &+\ \Pr\left[0\ chosen \mid 1\ optimal\right]\left(V_i^1 - V_i^0\right) \end{aligned} \tag{4}$$

Let the full (true) utility-difference function be

$$\begin{aligned} V_i^1 - V_i^0 &= \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \\ &= x_i^{'}\beta + \varepsilon_i \\ &= x_{-ki}^{'}\beta_{-k} + x_{ki}\beta_k + \varepsilon_i \end{aligned} \tag{5}$$

where each attribute-difference term $x_{ki} = X_{ki}^1 - X_{ki}^0$ concerns a single attribute *k* and its associated single indirect utility-difference coefficient, β_k. In a linear and additively separable model, this coefficient will be the same as the marginal utility of X_k.

The second line of equation (5) illustrates our convention for referring to the complete inner product, $x_i^{'}\beta$, of all attribute differences and their associated coefficients that actually enter into the systematic portion of the individual's utility function. The third line of the equation shows how we decompose this inner product into two terms, one being the inner product of all attribute differences other than x_{ki} and their corresponding parameters, denoted $x_{-ki}^{'}\beta_{-k}$, and the other being the k^{th} attribute difference and its own coefficient, $x_{ki}\beta_k$.

The probability that alternative 1 or alternative 0 is truly optimal for the individual (based on a full consideration of all attributes and their differences) would be given by

$$\begin{aligned} \Pr\left(1\ optimal\right) &= \Pr\left[x_i^{'}\beta + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_i^{'}\beta\right] \\ \Pr\left(0\ optimal\right) &= \Pr\left[\varepsilon_i > x_i^{'}\beta\right] \end{aligned} \tag{6}$$

In contrast, if the individual completely ignores attribute x_k (either because he or she does not think or bother to consider it, or believes incorrectly that it confers zero marginal utility), the probabilities of the observed choices will depend only on the levels of the other attributes, the vector x_{-k}:

$$\begin{aligned} \Pr\left(1\ chosen\right) &= \Pr\left[x_{-ki}^{'}\beta_{-k} + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_{-ki}^{'}\beta_{-k}\right] \\ \Pr\left(0\ chosen\right) &= \Pr\left[\varepsilon_i > x_{-ki}^{'}\beta_{-k}\right] \end{aligned} \tag{7}$$

There are thus two ways for the individual to make a "mistake." The subject could choose alternative 1 when alternative 0 is optimal, or choose 0 when 1 is optimal.

$$\begin{aligned}\Pr\left(1\ optimal \cap 0\ chosen\right) &= \Pr\left[\left(\varepsilon_i < x_i'\beta\right) \cap \left(\varepsilon_i > x_{-ki}'\beta_{-k}\right)\right] \\ \Pr\left(0\ optimal \cap 1\ chosen\right) &= \Pr\left[\left(\varepsilon_i > x_i'\beta\right) \cap \left(\varepsilon_i < x_{-ki}'\beta_{-k}\right)\right]\end{aligned} \tag{8}$$

It is a crucial assumption that we are talking about the same random error, ε_i, for each inequality. This is the portion of utility that is unobserved by the investigator, but known to the individual. It is typically assumed to account for all other unspecified attributes of each alternative, other than x_{1i} through x_{ki}, perhaps interacted with respondent characteristics, which drive the individual's choice. If the same error term is involved in each case in the probability formulas in (8), then the intersection of each pair of events is sometimes empty, depending upon the sign of $x_{ki}\beta_k$. Recall that $x_{-ki}'\beta_{-k} + x_{ki}\beta_k$ is just $x_i'\beta$, the systematic portion of the true indirect utility-difference. Thus $\Pr\left[\left(\varepsilon_i < x_i'\beta\right) \cap \left(\varepsilon_i > x_{-ki}'\beta_{-k}\right)\right]$ is nonzero only when $x_{ki}\beta_k$ is positive, and $\Pr\left[\left(\varepsilon_i > x_i'\beta\right) \cap \left(\varepsilon_i < x_{-ki}'\beta_{-k}\right)\right]$ is nonzero only when $x_{ki}\beta_k$ is negative.

The utility loss from the wrong choice due to ignoring x_{ki} is equal to $\left|V^1 - V^0\right|$, but the magnitude of this true utility difference is assumed to be unknown, ex ante. For any given ex post utility difference, the expected utility loss will be given by the probability of each type of mistake, times the resulting utility loss if such a mistake is made. Substituting these probabilities into the expression for the expected utility loss due to a wrong choice, $E\left[U\ Loss\right]$, yields the formula:

$$\begin{aligned}E\left[U\ Loss\right] &= \left[F\left(x_i'\beta\right) - F\left(x_{-ki}'\beta_{-k}\right)\right]\left(V^1 - V^0\right) \\ &\quad + \left[F\left(x_{-ki}'\beta_{-k}\right) - F\left(x_i'\beta\right)\right]\left(V^0 - V^1\right) \\ &= \left[F\left(x_{-ki}'\beta_{-k} + x_{ki}\beta_k\right) - F\left(x_{-ki}'\beta_{-k}\right)\right]\left(V^1 - V^0\right) \\ &\quad + \left[F\left(x_{-ki}'\beta_{-k}\right) - F\left(x_{-ki}'\beta_{-k} + x_{ki}\beta_k\right)\right]\left(V^0 - V^1\right)\end{aligned} \tag{9}$$

Either $\left(x_{ki}\beta_k\right)$ and $\left[F\left(x_{-ki}'\beta_{-k} + x_{ki}\beta_k\right) - F\left(x_{-ki}'\beta_{-k}\right)\right]$ are both positive, or they are both negative. Therefore, the utility difference can be equivalently expressed in terms of the absolute values of both terms.

$$E\left[U\ Loss\right] = 2 \times \left|F\left(x_{-ki}'\beta_{-k} + x_{ki}\beta_k\right) - F\left(x_{-ki}'\beta_{-k}\right)\right| \left|V^0 - V^1\right| \tag{10}$$

In words, the lost utility from an incorrect choice depends upon the sizes of the two absolute value terms in equation (10). The magnitude of the true utility-difference, $\left|V^0 - V^1\right|$ is unknown to the individual, ex ante. The first absolute value term, however, depends on two underlying quantities. We will call $\left|x_{-ki}'\beta_{-k}\right|$ the "other-attribute utility difference" and $\left|x_{ki}\beta_k\right|$ the "own-attribute utility difference." The argument inside the first absolute value

term is the cumulative density between the two limits of an interval of the underlying random variable. The interval is anchored at $x'_{-ki}\beta_{-k}$, the "other-attribute utility difference," and has a width given by $x_{ki}\beta_k$, the "own-attribute utility difference." It is simplest to address these two terms in reverse order.

2.1.2 Own-attribute utility differences: $x_{ki}\beta_k$

This quantity describes the *width* of the interval of the distribution of ε over which the probability of making an incorrect choice is calculated. The different utility contributions due to the k^{th} attribute itself, $\left|x_{ki}\beta_k\right|$, through its contribution to the expected utility loss from an incorrect choice, will affect the propensity to attend to the k^{th} attribute. The term $\left|x_{ki}\beta_k\right|$ can be large either if x_{ki} is large in absolute magnitude or if β_k is large in absolute magnitude. We thus expect that the propensity to attend will be greater, the greater the (positive or negative) contribution of this attribute to overall utility levels. If an attribute does not differ at all across alternatives, it should get little attention in the choice process. If the difference in the level of an attribute across alternatives (the range of levels for this attribute) is larger, then there is a greater likelihood that the subject will take this attribute into account in choosing between the two alternatives, ceteris paribus. Likewise, the greater the true marginal utility associated with this attribute, the greater the likelihood that it will receive more attention.

2.1.3 Other-attribute utility differences: $x'_{-ki}\beta_{-k}$

This quantity describes the anchoring point for the interval of the distribution of ε over which the probability of making an incorrect choice is calculated. For a given value of $\left|x_{ki}\beta_k\right|$, it will be the case that $\left|F\left(x'_{-ki}\beta_{-k}+x_{ki}\beta_k\right)-F\left(x'_{-ki}\beta_{-k}\right)\right|$, will be larger as the amount of cumulative density in this given-width interval of the distribution of ε_i is larger. In a typical random utility binary choice model, the error term is assumed to be standard logistic, and thus roughly bell-shaped and symmetric around zero. The cumulative density for such a distribution is thus greater when the interval is located near the middle of the distribution, rather than out in either tail. In general, then, the probability in question will be larger as $x'_{-ki}\beta_{-k}$ lies nearer to zero—namely, the more similar in utility levels are the alternatives in terms of all attributes other than the k^{th} attribute.

We thus expect that the propensity to attend to the k^{th} attribute will be greater, the closer $x'_{-ki}\beta_{-k}$ is to zero—in words, when the utility-difference based on all but attribute X_k is very close to zero. This will happen when (a.) all of the individual attribute-differences comprising x_{-ki} are close to zero, or (b.) the levels of the attribute differences that make up x_{-ki} vary in an offsetting or compensatory fashion within the individual's utility function. In words, benefits from paying attention to attribute *k* should be greatest when the choice based on the subset of attributes excluding attribute *k* is least clear.

2.2 Marginal Costs of Attention to an Additional Attribute

The theoretical model in the last section concerns the benefits from attention to an additional attribute. Attention should be devoted to a particular attribute as long as the expected marginal benefits of additional attention exceed the marginal costs of that additional attention. Certainly, the greater the marginal costs of considering an attribute, the larger must be those marginal benefits to induce the individual to pay attention to attribute *k*. The marginal (opportunity) costs to the individual of devoting attention to any incremental attribute *k* will depend on the individual's endowment of cognitive resources and how many other demands on those resources are currently active. There will be different demands on this cognitive capacity in any given time period. The forgone benefits from using this capacity elsewhere will define the opportunity cost of using it in the choice task in question. While we cannot directly measure the marginal opportunity costs of attention to an additional attribute in the illustrative example we will use here, we can hypothesize a number of properties that these cases should exhibit.

Suppose that the individual's finite cognitive resources can be allocated either to improving the accuracy of the choice at hand, or to other tasks. Cognitive resources are likely to be heterogeneous, so that the first few units of attention to the choice task at hand will cost relatively little in terms of forgone ability to concentrate on other tasks. However, the marginal opportunity cost of attention to an additional attribute can be expected to increase in the number of attributes being considered. This implies that the greater the number of attributes in a choice scenario, the higher the marginal cost of attention to any specific attribute. When the same number of attributes is used for all choice sets in a study, of course, the size of this effect cannot be estimated.

Anything that improves the subject's ability to convert a given amount of cognitive resources into choice accuracy for the designated choice task (the "technology" for producing accurate choices) may uniformly reduce the opportunity cost of attention or "cognitive price" for attention to an additional attribute. These factors could include advance preparation or training for the choice exercise, or previous experience with similar choice tasks.[7]

Qualitatively, the number of other cognitive tasks across which the individual must allocate scarce cognitive resources can affect the optimal allocation of the resources to the choice task in question. It would be helpful to control for differences along this dimension by eliciting information about the extent to which the individual is preoccupied by other cognitive challenges.[8]

Factors which affect the marginal costs of attention to various attributes to different extents will be particularly important to consider. For example, consider the strategy, by

[7] For example, if the subject is a member of a regular panel, there will be records of the number of similar choice experiments in which she has participated. Conversely, anything that makes it harder to do other cognitive tasks could affect the relative price of attention to the current choice task. For example, controlling for cognitive capacity, if a respondent is housebound or neither at school nor in the labour force, he may be inclined to devote a greater share of his cognitive capacity to the choice exercise in question, rather than to other tasks.

[8] For student samples, this could include a measure of how crowded the student's calendar is (e.g. how many units they are currently taking or how many hours they work, or even how many hours of homework they already know that they must complete on the day of the choice experiment). Or, we might suspect that cognitive resources allocated to the choice task at hand are partially an artefact of other demands on the respondent's attention. Debriefing questions might attempt to elicit categorical or open-ended information concerning other problems that may be competing for the subject's scarce attention at the time of the choice exercise.

some advertisers, to put important information about products in the "fine print" of their offers. Or, consider national-level advertisements which increase the cost of attention to the price attribute by merely advising the potential customer to "contact your local dealer for price information." The order of a particular attribute in a long list of product attributes (at the top, versus the middle or the bottom) may also affect the cost of paying attention to that attribute.

We mention considerations about differing marginal costs of attention to different attributes, at this point in the paper, for completeness. In the empirical example we offer in this paper, we do not have sufficient data (i.e. measured variability) in respondents' marginal costs of attention to our different attributes to permit us to control for these types of differences during estimation. Any attention-cost variables in our inventory will affect the costs of attention to all attributes, and in most cases these are fixed across choice sets for each respondent. If these costs affect attention to all attributes proportionally, the outcome should be observationally equivalent to allowing for different scale factors for each individual's choices (with potentially no effect on the relative sizes of the estimated marginal utilities). It will be important for future studies to collect sufficient information to explore the influence of differing marginal costs of attention to different attributes and the effects of these differences on estimated marginal utilities (and hence, potentially, on measures such as WTP). In any event, it is these (implicit) marginal costs of attention that will restrict the individual's attention to attributes to be less than that amount which would correspond to driving the marginal benefits of attention all the way to zero.

2.3 Multiple-Alternative Cases

2.3.1 Generalized other-attribute effects

In the two-alternative case, we could make a strong case that the closeness in utility space of the two alternatives on the basis of their other attributes should increase the attention paid to each additional attribute. With just two alternatives, the simple absolute difference in systematic utilities according to other attributes, $\left|x_{-ki}'\beta_{-k}\right|$, will adequately capture the relevant properties of the choice set. If differentiability proves helpful in whatever estimation method is used, this measure could be proxied by $\left(x_{-ki}'\beta_{-k}\right)^2$.

With three or more alternatives, however, we need an analogous, but more general proxy for this information. We still wish to capture the extent to which there is a clear-cut "best" option among the available alternatives, based on all attributes other than this one. One possible measure would be the difference between the two leading alternatives, based on all attributes other than the one in question. We will call this statistic $lead\left(x_{-ki}'\beta_{-k}\right)$, the side of the lead, in utility units, enjoyed by the front-running alternative based on all other attributes. This statistic amounts to the absolute difference between the first two order statistics across alternatives for the other-attribute utility levels. In the three-alternative case, the researcher would need to calculate each of the indirect utility differences, relative to the numeraire third alternative: $\left(x_{-ki}^{1'}\beta_{-k}\right)$, $\left(x_{-ki}^{2'}\beta_{-k}\right)$ and 0. The two largest-valued differences would need to be identified (i.e. the largest value and the median value), and their absolute difference calculated. For any current set of estimates for the true marginal utility parameters, β, this computation is tractable, but there is a risk that it will be awkward in a maximum likelihood context since this choice set statistic is not smoothly differentiable.

Other proxies can also be considered, however. The researcher can calculate the indirect utility differences, $\left(x^{1'}_{-ki}\beta_{-k}\right)$, $\left(x^{2'}_{-ki}\beta_{-k}\right)$ and 0, and compute the standard deviation in these quantities. The greater the standard deviation in these measures, the more "different" are the alternatives in terms of all other attributes. We will denote this measure as $sd\left(x'_{-ki}\beta_{-k}\right)$.

It may also be relevant to compute the degree of skewness in these quantities across alternatives. In the three-alternative case considered in our empirical section, the more positively skewed is this distribution of three systematic net utilities, the farther apart are the two highest values, relative to the lowest value. For a given standard deviation, greater positive skewness in the three-alternative case means there is more of a "clear winner" among the three alternatives in terms of "all but the k^{th} attribute." We will call this measure $skew\left(x'_{-ki}\beta_{-k}\right)$. However, the same degree of skewness can be present for triples of values which have very different standard deviations, so it will be necessary to control for standard deviation before using skewness to measure the extent of a positive outlier in any triple of values.

A fourth candidate for dissimilarity might be an entropy measure, such as that employed in Swait and Adamowicz (2001a,b). There is a roughly quadratic bound on entropy that varies systematically with the standard deviation, but the relationship is not exact (Duquette et al. 2009). All of these various measures of the dissimilarity of alternatives based on other attributes (directly related to the extent to which there is a clear winner based on the other attributes) will be referred to generically as $dissim\left(x'_{-ki}\beta_{-k}\right)$.

2.3.2 Generalized own-attribute effects

In the two-alternative case, the potential for the k^{th} attribute to change the identity of the winning alternative is also argued to be a logical determinant of attention to that k^{th} attribute. This potential will depend both on the marginal utility associated with that attribute and the extent to which the level of that attribute differs across alternatives. If either component is large, the attribute will attract more attention. In the two-alternative case, simply the absolute difference $\left|x_{ki}\beta_k\right|$ is sufficient to capture the combination of these two effects. If a smoothly differentiable proxy for this quantity is required, perhaps $\left(x_{ki}\beta_k\right)^2$ could be employed.

In the multi-alternative case, the various utility-contributions across alternatives of the k^{th} attribute, relative to the numeraire alternative, must be taken into account. These will be $x^1_{ki}\beta_k$, $x^2_{ki}\beta_k$, and 0. As for the other-attribute utility differences, we might postulate that attention to the k^{th} attribute should be increasing in size of the lead enjoyed by the alternative with the largest value for this utility contribution, which we will denote $lead\left(x_{ki}\beta_k\right)$. Alternatively, we might use the standard deviation across alternatives of the utility-contributions due the k^{th} attribute, $sd\left(x_{ki}\beta_k\right)$, and/or the skewness of the distribution (across alternatives) of these own-attribute effects: $skew\left(x_{ki}\beta_k\right)$.

These various measures of dissimilarity of alternatives based on the attribute in question (directly related to the ability of this attribute to make a big difference to the overall utility levels from each alternative) will be referred to generically as $dissim\left(x_{ki}\beta_k\right)$.

3 Empirical Models for Attention to Attributes

Needed is a strategy to incorporate into a choice model any systematic differences in expected benefits (and costs, where possible) from paying attention to particular attributes. The researcher does not know, ex ante, whether there may be attributes which (a.) are truly relevant to the subject's choice decision, but (b.) are more-or-less ignored because of the individual's benefit-cost assessment that reflects scarce cognitive capacity and the avoided expected losses from an incorrect decision. All that the investigator can determine is, on average, which attributes appear to be less relevant to choices and which attributes do seem to affect them.

3.1 Propensity to Attend

We introduce the idea of a multiplicative "propensity to attend" parameter, a_{ki}, associated with the k^{th} attribute. This propensity to attend can be viewed as affecting either the apparent marginal utility associated with the k^{th} attribute (by converting β_k to $\beta_k a_{ki}$) or, as the perceived difference in the level of this attribute across the two alternatives (by converting x_{ki} to $a_{ki}x_{ki}$). The indirect utility difference driving the basic conditional logit choice model in equation (2) can thus be recast as:

$$V_i^1 - V_i^0 = \sum_{k=1}^{K} \beta_k a_{ki} x_{ki} + \varepsilon_i \tag{11}$$

We will let a_{ki} vary systematically across individuals by allowing it to depend upon an "index" $Z_{ki}^{'}\gamma_k$ consisting of a vector of variables $Z_{ki}^{'}$ and additional parameters γ_k. There are certainly several potential strategies for incorporating systematically differing propensities to attend to different attributes in a choice set. Some of these might be:

$$\begin{aligned} V_i^1 - V_i^0 &= \sum_{k=1}^{K} \beta_k \left[1 + Z_{ki}^{'}\gamma_k\right] x_{ki} + \varepsilon_i, \text{ or} \\ V_i^1 - V_i^0 &= \sum_{k=1}^{K} \beta_k F\left[Z_{ki}^{'}\gamma_k\right] x_{ki} + \varepsilon_i, \text{ or} \\ V_i^1 - V_i^0 &= \sum_{k=1}^{K} \beta_k \exp\left(Z_{ki}^{'}\gamma_k\right) x_{ki} + \varepsilon_i \end{aligned} \tag{12}$$

The first formulation treats $\left[1 + Z_{ki}^{'}\gamma_k\right]$ directly as the propensity to attend without constraining the sign or size of this propensity factor. The second candidate formulation treats the index $Z_{ki}^{'}\gamma_k$ as a latent variable that can take on any value over the entire real line. However, if we use a function like an inverse log-odds transformation (a standard logistic cumulative density function) for *F*, we can restrict the effective value of the propensity factor to lie strictly between zero and one, with zero interpreted as "no attention" and one

interpreted as "complete attention."[9] The third formulation treats the propensity to attend to each attribute as simply a non-negative factor that scales the true marginal utility associated with an attribute either up or down as the value of this factor is greater or less than one. The sign of the underlying true marginal utility, β_k, is thereby preserved. This assumption may be the most empirically hospitable one when a sign restriction is desired.[10,11]

3.1.1 Implementation

Our theory section has suggested specific information that should be included among the Z_{ki} variables that determine the subject's propensity to attend to the k^{th} attribute of the alternatives in the choice set. These factors contribute to the expected benefits (or the costs) of paying attention to a particular attribute. The propensity-to-attend measure, a_{ki}, should be some explicit function of how different the alternatives are in terms of the utility based on *all other attributes*, which we will denote by the construct that measures this in the two-alternative case, $\left|x'_{-ki}\beta_{-k}\right|$. The propensity-to-attend measure will also depend on the difference across alternatives in utility derived from *this attribute*, denoted by $\left|x_{ki}\beta_k\right|$ for the two-alternative case. Finally, it will depend on any available variables which capture the *marginal cost* of attention to this attribute (which may or may not differ across attributes *k*).

Taking the first specification in equation (12) as our example, we now differentiate among the generic coefficients in $a_{ki} = \left[1 + Z'_{ki}\gamma_k\right]$, by distinguishing three types of parameters:

$$a_{ki} = 1 + \left(\alpha_k \left|x'_{-ki}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right| + C'_{ki}\delta_k\right) \tag{13}$$

[9] The slight inconvenience in estimating such a model stems from the starting values to be used. If all of the parameters γ_k are simultaneously zero, then $a_{ki} = F\left[Z'_{ki}\gamma_k\right] = 0.5$. The apparent marginal utilities from a naive random utility specification would therefore be obtained from the first model in equation (12) only if the starting values for the β_k parameters were all to be doubled. The default assumption (i.e. $\gamma_k = 0$) is therefore that the "true" marginal utilities are actually twice what they appear to be in the naive model. As the index $Z'_{ki}\gamma_k$ is larger, the true marginal utility will be less than twice its apparent naive value; as the index is smaller, the true marginal utility will be more than twice its apparent naive value. In the limit, as $Z'_{ki}\gamma_k$ goes to +∞, the implied propensity to attend goes to 1.0 and the associated β_k corresponds to the true marginal utility. The counterfactual of interest in this model corresponds to the question of what would have been the marginal utilities if the subject had been paying full attention to all attributes. The answers are contained in the estimates of each β_k from this specification.

[10] Certainly, if a nonlinear model is used, and if analytical derivatives are to be employed, this formulation would be easier than the inverse log-odds transformation suggested for the case where marginal propensities are constrained to lie on the 0,1 interval.

[11] Fortunately, the ratios of estimated marginal utilities are all that matter for welfare estimates, and the "true" marginal utilities can be known only up to a scale factor. The relevant counterfactual in this case again concerns what would be the size of the estimated marginal utilities if all attributes received *equal* attention. A logical value for this equal propensity to attend would be 1.0, which would also constitute a logical starting assumption, since $a_{ki} = \exp\left(Z'_{ki}\gamma_k\right) = 1$ if $Z'_{ki}\gamma_k = 0$, which will be the case if the vector of parameters γ_k is initially assumed to be a zero vector.

where the vector $C_{ki}^{'}$ is a set of variables, when available, that capture the individual's cognitive marginal costs of evaluating attribute *k*. If $\alpha_k = \theta_k = \delta_k = 0$, then the propensity to attend, a_{ki}, equals exactly one for all attributes, the desired case.

We do not necessarily expect the expressions that capture the benefits of attention $\left(\alpha_k \left|x_{-ki}^{'}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right|\right)$ or the costs of attention $\left(C_{ki}^{'}\delta_k\right)$ to be the same across attributes, because the constructed variables $\left|x_{-ki}^{'}\beta_{-k}\right|$, $\left|x_{ki}\beta_k\right|$, and possibly the relevant vector of variables $C_{ki}^{'}$ will differ across attributes. However, perhaps the incremental effects of the choice set design variables and individual characteristics that determine the net benefits of attention to attributes (i.e. the coefficients α_k and θ_k for benefits, or γ_k for costs) could be the same across attributes *k=1,...,K*, so that the corresponding coefficients can be constrained to be equal across attributes:

$$a_{ki} = 1 + (\alpha\left|x_{-ki}^{'}\beta_{-k}\right| + \theta\left|x_{ki}\beta_k\right| + C_{ki}^{'}\delta) \tag{14}$$

Where possible, one should estimate models with and without these restrictions and test whether the restrictions can be rejected. These types of restrictions are possible because all of the variables in question for α and θ are in utility-units, not the units of the raw attributes.

If the cost-of-attention variables do not differ quantifiably across attributes, so that $C_{ki}^{'} = C_i^{'}$, it may actually be necessary to constrain $\delta_k = \delta$. Otherwise, the attribute-specific parameters δ_k are likely to be difficult or impossible to identify separately from the marginal utilities, β_k. If we assume that the effects of the cost variables are the same across all attributes—e specially if we adopt the version of a_{ki} in the third line of equation (12)—then a_{ki} can be readily factored into two components: $\exp\left(\alpha_k \left|x_{-ki}^{'}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right|\right)$ and $\exp\left(C_{ki}^{'}\delta_k\right)$, and the implied form of the indirect utility difference will be:

$$\begin{aligned} V_i^1 - V_i^0 &= \sum_{k=1}^{K} \beta_k \exp\left(\alpha_k \left|x_{-ki}^{'}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right|\right) \exp\left(C_i^{'}\delta\right) x_{ki} + \varepsilon_i \\ &= \exp\left(C_i^{'}\delta\right) \sum_{k=1}^{K} \beta_k \exp\left(\alpha_k \left|x_{-ki}^{'}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right|\right) x_{ki} + \varepsilon_i \end{aligned} \tag{15}$$

Since $\exp\left(C_i^{'}\delta\right)$ is strictly positive and utility is invariant to the scale of measurement, we could divide through by $\exp\left(C_i^{'}\delta\right)$ to produce a heteroskedastic model:

$$V_i^1 - V_i^0 = \sum_{k=1}^{K} \beta_k \exp\left(\alpha_k \left|x_{-ki}^{'}\beta_{-k}\right| + \theta_k \left|x_{ki}\beta_k\right|\right) x_{ki} + \frac{\varepsilon_i}{\exp\left(C_i^{'}\delta\right)} \tag{16}$$

In the model with a strictly positive propensity to attend to attributes, the variables $C_i^{'}$ that capture cognitive cost differences across individuals can enter the model equivalently as factors that affect the dispersion in the conditional logit error term. Any variable that

increases the overall cost of attention should tend to decrease the respondent's propensity to attend to every attribute. Lesser attention can be expected to increase the error variance in the model.

4 Empirical Example

4.1 A stated preference survey concerning morbidity/mortality risk reduction programs

As a simple illustration, we use choice data from a stated preference survey concerning individuals' preferences over health-risk reduction programs. Cameron and DeShazo (2009) use SP methods to elicit preferences for programs to reduce the risk of morbidity and mortality in a general-population sample of adults in the United States. The survey was fielded by Knowledge Networks, Inc., to their standing consumer panel, using a combination of internet and WebTV interfaces. The details of the survey and its basic analysis have been documented very extensively elsewhere, so we do not repeat the information here.[12]

In brief, the Cameron/DeShazo survey consists of five modules. We will outline these models only to explain the sources of some of the covariates we will use in our illustration. The first module asks respondents, among a variety of other questions, to rate their subjective risks, from low (-2) to high (+2), of contracting each of a range of major illnesses or injuries.

The second module is a tutorial that explains the concept of an "illness profile." This is a description of a sequence of future health states associated with a specified major illness or injury that the respondent may face over his or her remaining lifetime. An illness profile includes the years before the individual becomes sick (latency), illness-years while the individual is sick, recovered/post-illness-years after the individual more-or-less recovers from the illness, and lost life-years if the individual dies earlier than he would have in the absence of the illness or injury. After the tutorial about illness profiles, the individual is informed that he might be able to purchase new programs that would reduce his risk of experiencing certain illness profiles. Each illness-related risk-reduction program described in the survey consists of diagnostic blood tests, drug therapies, and life-style changes, and would be available at a specified annual cost, to be paid on a recurring basis as long as the individual is neither sick with this illness nor dead.

The key module of each survey involves a set of five different three-alternative conjoint choice tasks where the individual is asked to choose one of two possible health-risk reducing programs or a status quo alternative. Each program reduces the individual's risk of experiencing the corresponding illness profile. The illness profiles are described succinctly for each of the choice tasks—in terms of the baseline probability, age at onset, duration, and eventual outcome (recovery or death). Each corresponding risk reduction program is defined in terms of the extent to which it can be expected to reduce this risk, and its monthly and annual cost. Figure 1 provides one instance of the type of a stated choice scenario posed to respondents.

[12] 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 Networks Panel and Sample Selection Corrections, Appendix E - Model, Estimation and Alternative Analyses, and Appendix F - Estimating Sample Codebook.

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 1: An Example of a Choice Set Summary Table

Module 4 contains debriefing questions to cross-check the internal consistency of responses. Module 5 is collected separately from our survey and contains detailed socio-demographic data for the individual and their household, as well as responses to a battery of health-related questions (including any illnesses the individual has already faced).

Table 1 contains descriptive statistics for the empirical models to be used in this paper. It summarizes only those variables pertinent to the present illustration. These include raw data, but this information is processed before use, based on the economic theory of discounted expected utility, to yield the necessary constructed variables for our analysis to be discussed below. Finally, in some of our specifications, we allow for preferences to differ systematically with exogenous characteristics of the respondent (gender and age). We also allow preferences to differ according to the individual's subjective risk rating for the illness/injury in question, with the average of their subjective risk ratings for all of the other major risk categories covered by our survey, and with any individual subjective adjustments of the surveys' statements about the existence of likely benefits and the latency of the health risk.

Table 1: Descriptive Statistics for Variables Used in Estimating Specifications

Variable	Description	Mean	Std. Dev.	Min	Max
Program attributes (14074 programs)					
- Raw illness/program attributes					
Cost	Annual cost of program (paid when not sick or dead)	355.00	341.14	24	1680
$\Delta\Pi_i^{AS}$	Risk change (i.e. negative, a risk reduction)	-0.0034	0.0017	-0.006	-0.001
Latency	Years until illness/injury begins	19.65	12.03	1	60
Sick years	Duration of illness/injury (years)	6.53	7.21	0	52
Recovered years	Number of years in post-illness health state	1.62	4.62	0	55
Lost life-years	Number of life-years lost	10.87	10.32	0	55
- Constructed variables					
(income term)	Net income under each alternative	-0.052747	0.048772	-0.2513	0.1083
$\Delta\Pi_i^{AS}$ log(pdvi+1)	Term in present discounted sick-years	-0.003111	0.003006	-0.01710	0
$\Delta\Pi_i^{AS}$ log(pdvr+1)	Term in present discounted recovered-years	-0.003374	0.003189	-0.01711	0
$\Delta\Pi_i^{AS}$ log(pdvl+1)	Term in present discounted lost life-years	-0.000746	0.001841	-0.01648	0
Sasubrsk (mean = msasubrsk)	Same-illness subjective risk rating (-2 = low, 2=high)	-0.2593	1.2531	-2	2
Cosubrsk (mean = mcosubrsk)	Average subjective risk rating (other major health risks)	-0.2537	0.8670	-2	2
(benefits never)	=1 if expects never to benefit from this program	0.0759	0.2648	0	1
(min overest latency)	Minimum overestimate of the latency of the health risk	-7.483	11.98	-58	29
Respondent characteristics (1519 respondents)					
Income	Annual income (dollars)	51048	33781	5000	150000
Female	=1 if female	0.5135	0.5000	0	1
age (mean = mage)	Age in years at time of response	50.11	15.18	25	93

4.2 Estimating specification for the naïve choice model

We will use a simplified version of the theoretical model presented by Cameron and DeShazo (2009). In that paper, it is established that stated choices in this general population sample appear to be best predicted by a model that involves discounted expected utility from durations in different adverse future health states. Here, we will outline the model only briefly, to justify why the variables which explain choices are constructed in the ways that they are. The choice scenarios involve probabilistic sequences of future events, so a model based on discounted expected utility is about the simplest reasonable starting point.

To understand the basic model, consider just a pair-wise choice between Program A and the status-quo alternative (N). Define the discount rate as r and let $\delta^t = (1+r)^{-t}$. For individual i, let Π_i^{NS} be the probability of suffering a given adverse health profile (i.e. getting "sick") if the status quo alternative is selected, and let Π_i^{AS} be the (reduced) probability of suffering this adverse health profile if Program A is chosen. Thus $\Delta\Pi_i^{AS} = \Pi_i^{NS} - \Pi_i^{AS}$ is negative, since this is the risk reduction to be achieved by Program A.

The sequence of health states that makes up the illness profile to be addressed by Program A is captured by a set of mutually exclusive and exhaustive (0,1) indicator variables associated with each future time period, T_i: $1\left(\textit{pre-illness}_{it}^A\right)$ for years in the latency period prior to any symptoms, $1\left(\textit{illness}_{it}^A\right)$ for illness-years, $1\left(\textit{recovered}_{it}^A\right)$ for recovered, remission, or post-illness years, and $1\left(\textit{life-year lost}_{it}^A\right)$ for a year of premature mortality. Individuals are modelled as expecting to pay the annual cost of the risk reduction program only if they are neither sick nor dead.

The algebra of calculating present discounted expected utility differences is simplified considerably because we model health states as being uniform within specified intervals, as are income and program costs in this model. This feature allows us to discount health states first, and then take expectations. The present discounted number of years making up the remainder of the individual's nominal life expectancy, T_i, is given by $pdvc_i = \sum_{t=1}^{T_i} \delta^t$. Other relevant discounted spells, also summed from $t=1$ to T_i include: $pdve_{it}^A = \sum \delta^t 1\left(\textit{pre-illness}_{it}^A\right)$, $pdvi_{it}^A = \sum \delta^t 1\left(\textit{illness}_{it}^A\right)$, $pdvr_{it}^A = \sum \delta^t 1\left(\textit{recovered}_{it}^A\right)$, and $pdvl_{it}^A = \sum \delta^t 1\left(\textit{life-year lost}_{it}^A\right)$.

Since the different health states exhaust the individual's nominal life expectancy, $pdve_i^A + pdvi_i^A + pdvr_i^A + pdvl_i^A = pdvc_i$. Finally, to accommodate the assumption that each individual expects to pay program costs only during the pre-illness or recovered post-illness periods, $pdvp_i^A = pdve_i^A + pdvr_i^A$ is defined as the present discounted (healthy) time over which payments must be made. This can be interpreted as the expected discounted duration of program costs, with the expectation taken across whether or not the individual gets sick.

To further simplify notation, let $cterm_i^A = \left[\left(1-\Pi_i^{AS}\right) pdvc_i + \Pi_i^{AS}\, pdvp_i^A\right]$ and let $yterm_i^A = \left[\left(1-\Pi_i^{NS}\right) pdvc_i + \Pi_i^{NS}\, pdvp_i^A\right] - \Delta\Pi_i^{AS}\, pdvi_i^A$. These two terms account for the

pattern of income net of program costs over time as a function of probabilistic health states. Then the expected utility-difference that drives the individual's choice between Program A and the status quo can be defined as follows (where there will be an analogous term for the utility difference between Program B and the status quo in our three-alternative model):

$$\begin{aligned}\Delta PDV\left(E\left[V_i^A\right]\right) &= \beta_1\left\{\left(Y_i - c_i^A\right) cterm_i^A + Y_i\, yterm_i^A\right\} \\ &+ \beta_2\left\{\Delta\Pi_i^A\, pdvi_i^A\right\} + \beta_3\left\{\Delta\Pi_i^A\, pdvr_i^A\right\} + \beta_4\left\{\Delta\Pi_i^A\, pdvl_i^A\right\} + \varepsilon_i^A\end{aligned} \tag{17}$$

The four terms in braces can be constructed from the data, given specific assumptions about the discount rate.[13] In this application, these constructed variables are the x_{ki} (the differences in the attribute levels between each substantive alternative and the status quo).

The empirical results described in Cameron and DeShazo (2009) suggest that a basic four-parameter, homogeneous-preferences model such as that in equation (17) is dominated by a specification that is not merely linear in the terms involving present discounted health-state years. Factoring out the probability difference from the final substantive term in equation (17) gives:

$$\begin{aligned}&\beta_2\left\{\Delta\Pi_i^j\, pdvi_i^j\right\} + \beta_3\left\{\Delta\Pi_i^j\, pdvr_i^j\right\} + \beta_4\left\{\Delta\Pi_i^j\, pdvl_i^j\right\} \\ &\quad = \Delta\Pi_i^j\left[\beta_2\, pdvi_i^j + \beta_3\, pdvr_i^j + \beta_4\, pdvl_i^j\right]\end{aligned} \tag{18}$$

where *j*=*A*,*B*,*N*, and $pdvX_i^N = 0$ for $X = i, r, l$ since the durations in adverse health states are all normalized to zero for the numeraire status quo alternative. However, this simple linear specification does not explain respondents' observed choices as well as a model that employs shifted logarithms of the $pdvX_i^j$ terms:

$$\Delta\Pi_i^j\left[\beta_2 \log\left(pdvi_i^j + 1\right) + \beta_3 \log\left(pdvr_i^j + 1\right) + \beta_4 \log\left(pdvl_i^j + 1\right)\right] \tag{19}$$

The basic discounted expected utility-difference specification that is presumed to drive respondent's choices is therefore:

[13] In this paper, we assume a common discount rate of five percent. In Cameron and DeShazo (2009), we explore the consequences of assuming either a three percent or a seven percent alternative discount rate. Work in progress involves the estimation of individual-specific discount rates simultaneously with these stated choices concerning health risk reduction programs, using additional data on inter-temporal choices by a separate sample of respondents from the same population.

$$\begin{aligned}\Delta PDV\left(E\left[V_i^j\right]\right) &= \beta_1\left\{\left(Y_i - c_i^j\right)cterm_i^j + Y_i yterm_i^j\right\} \\ &+ \beta_2\left\{\Delta\Pi_i^j \log\left(pdvi_i^j + 1\right)\right\} \\ &+ \beta_3\left\{\Delta\Pi_i^j \log\left(pdvr_i^j + 1\right)\right\} \\ &+ \beta_4\left\{\Delta\Pi_i^j \log\left(pdvl_i^j + 1\right)\right\} + \varepsilon_i^j \\ &= x_i^j\beta + \varepsilon_i^j\end{aligned} \tag{20}$$

There is an analogous term for Program B in the three-way choice context.

In the empirical estimates that follow, our "basic linear" model involves these four constructed variables in the sets of braces in equation (20), and their four estimated parameters, $\left(\beta_1, \beta_2, \beta_3, \beta_4\right)$. We assume that a researcher who ignores the effects of scenario design (specifically, the mix of attribute levels presented in a choice set) would merely estimate this simple model. In Cameron and DeShazo (2009) we show how the estimated model can be used to build estimates of willingness to pay for a microrisk reduction (10^{-6}) in the chance of suffering from a specific type of illness profile. This construct is essentially a generalization of the more-restrictive concept of the value of a statistical life (*VSL*) commonly employed in the mortality risk valuation literature. For this paper, however, we concentrate mainly on the estimation of the four parameters in equation (20) and the extent to which attention to these four different attributes may be biased as a result of the design of our choice sets.

4.3 Potential attention biases: measurement and control

To construct measures for the similarity of alternatives based on all attributes other than the one in question, it is necessary to have measures of the "true" marginal utilities of each attribute, uncontaminated by attention biases. To identify these true marginal utilities, however, it is necessary to control for attention biases.

Ideally, one would specify a conditional logit choice model where each additively separable marginal utility parameter is allowed to shift with variables which measure $dissim\left(x'_{-ki}\beta_{-k}\right)$ and $dissim\left(x_{ki}\beta_k\right)$, for each attribute. However, these dissimilarity variables will each be a fairly complicated function of the same basic vector of "true" marginal utility parameters that they modify. It is straightforward (if tedious) to write down the log-likelihood for full information maximum likelihood estimation of this model (using any of the practical candidates for these dissimilarity measures). However, given the complex and repeated manner in which the basic utility parameters enter the model, one can expect the log-likelihood function to be somewhat difficult to maximize.

To allow us to explore these data for evidence of unequal attention bias across attributes, however, it is possible to implement a crude correction without resorting to custom-programmed nonlinear optimization models. Estimation can be accomplished by employing an iterative algorithm that relies solely on sequential utilization of packaged conditional logit algorithms. This iterative algorithm is described in detail in Appendix A (this seems to mimic the method used by Swait and Adamowicz (2001a,b) in their work with entropy as a measure of choice set complexity). Upon convergence, the last set of parameters can be used to compute the "final" estimated values of the shift variables capturing, for each attribute, the similarity of the available alternatives based on the other attributes, and the dissimilarity of the available alternatives based on this attribute. In our

model with four marginal utility parameters, there will be eight additional (estimated) regressors to be interacted with the basic attribute variables.

When the model is estimated iteratively in this fashion, using packaged conditional logit software, the parameter variance-covariance matrix in the last round, of course, does not reflect the estimated nature of the estimated other- and own-attribute standard deviations (or "leads") in utility. Full information maximum likelihood estimation is required to estimate all of the parameters of the two models simultaneously, so that a full parameter variance-covariance matrix can be obtained.[14]

4.4 Empirical results

In this example, we expect to find a positive marginal utility of income (β_1), and negative marginal utilities associated with the logarithms of (shifted) present discounted sick-years, recovered-years, and lost life-years $(\beta_2, \beta_3, \beta_4)$. We expect that the greater the disparity in utilities across alternatives, based on other attributes, the less will be the individual's apparent responsiveness to differences in the level of any given attribute. We also expect that the greater the difference in utility derived from the attribute in question, the greater will be the individual's apparent responsiveness to differences in the level of any given attribute.

4.4.1 Models using $sd\left(x_{-ki}'\beta_{-k}\right)$ and $sd\left(x_{ki}\beta_k\right)$:

Table 2 shows the results of a succession of fixed-effects conditional logit-type models where the disparities in indirect utility based on all other attributes, and based on just this attribute, are measured as the standard deviation across alternatives. Model SD1 (homogeneous preferences) is a baseline specification, with no attention-related correction terms, involving only the four utility parameters in our most basic specification. The signs on all three estimated parameters are as anticipated, and each is strongly statistically significantly different from zero.

Model SD2 (heterogeneous preferences) generalizes this specification to allow for systematically varying preference parameters. The marginal utility of net income is statistically significantly higher for women. The coefficient capturing the marginal disutility of expected discounted sick-years is negative. It is more negative, the higher the individual's subjective risk of suffering the illness or injury targeted by the risk-reduction program in question. It is less negative, the higher the individual's average subjective risk of suffering from any of the other major categories of health risks addressed in the survey. If the individual indicates, ex post, that they expect never to benefit from the program in question, the disutility from illness in this case is drastically reduced. Finally, the greater the individual's overestimate of the latency period before benefits begin (i.e. before the illness will cause pain or disability), the lesser the implied disutility from present discounted sick-time.

If recovered-years are viewed as a return to perfect health, we would expect utility in that state should be identical to pre-illness utility, but our estimates suggest that most

[14] We have explored a number of FIML specifications for the overall optimization process, using Matlab's general function-optimizing software. As would be expected, however, it can be very difficult to achieve convergence in this context because the various utility parameters in the model enter multiplicatively. One can expect the iterated estimates used in the body of this paper to understate the amount of noise in the estimates, to a degree, because the dissimilarity variables are treated as non-stochastic when they are actually estimated quantities.

individuals do not view this to be the case. There appears to be negative utility associated with "recovered" years, and this disutility is greater, the older the respondent at the time when these stated choices are being made. These are major illnesses, including five types of cancers, heart disease or heart attack, respiratory disease, stroke, diabetes, Alzheimer's disease and traffic accidents. An expectation of lingering morbidity is reasonable.

Table 2: "Standard Deviation" Variant: Uncorrected and Attention-corrected **Fixed-Effects** Conditional Logit Models for Health-Risk Reduction Programs (with homogeneous and heterogeneous preferences)

	Exp. sign	(SD1) Homogeneous preferences	(SD2) Heterogeneous preference	(SD3) Attention homogenous-homogeneous	(SD4) Attention heterogeneous-heterogeneous	(SD5) Attention heterogeneous-homogeneous	(Lead5) Attention heterogeneous-homogeneous
Income term (β_1)							
(income term)	**[+]**	3.148	2.941	2.759	3.455	1.514	2.685
		(7.77)***	(5.16)***	(3.96)***	(5.89)***	(2.93)***	(5.82)***
... × (sd(U othr attr)-mean sd)	**[-]**	-	-	7.715	3.152	-1.505	-.4693
				(2.86)***	(3.11)***	(1.78)*	(0.74)
... × (sd(U this attr)-mean sd)	**[+]**	-	-	4.882	-3.507	2.361	1.545
				(1.08)	(1.92)*	(1.89)*	(1.38)
…× female		-	3.916	-	5.473	-	-
			(5.58)***		(5.30)***		
Sick-years term (β_2)							
$\Delta\Pi_i^{AS}$ log(pdvi+1)	**[-]**	-27.06	-14.39	-23.61	-20.57	-7.124	-11.23
		(4.50)***	(1.93)*	(2.46)**	(2.65)***	(1.02)	(1.69)*
... × (sd(U othr attr)-mean sd)	**[+]**	-	-	-48.13	-21.01	91.18	63.95
				(1.04)	(1.05)	(5.36)***	(4.24)***
... × (sd(U this attr)-mean sd)	**[-]**	-	-	-4.894	21.08	-106.1	-136.8
				(0.04)	(0.92)	(6.48)***	(10.94)***
... × (sasubrsk-msasubrsk)		-	-22.09	-	-26.32	-	-
			(3.80)***		(4.32)***		
... × (cosubrsk-mcosubrsk)		-	27.44	-	31.64	-	-
			(3.20)***		(3.61)***		
... × (benefits never)		-	137.4	-	130	-	-
			(4.14)***		(3.82)***		
... × (min overest latency)		-	8.13	-	9.24	-	-
			(12.53)***		(11.88)***		
Recovered-years term (β_3)							
$\Delta\Pi_i^{AS}$ log(pdvr+1)	**[-]**	-24.03	-40.55	-72.19	-43.81	-33.92	-45.3
		(2.51)**	(3.98)***	(3.69)***	(2.28)**	(2.16)**	(2.99)***

... × (sd(U othr attr)-mean sd)	**[+]**	-		-57.4	.3823	31.98	39.68
				(0.51)	(0.01)	(1.49)	(2.12)**
... × (sd(U this attr)-mean sd)	**[-]**	-		122.3	13.45	56.41	32.27
				(1.92)*	(0.13)	(0.68)	(0.58)
... × (age-mage)		-	-1.305	-	-1.348	-	-
			(1.95)*		(1.62)		
Lost life-years term (β_4)							
$\Delta\Pi_i^{AS}$ log(pdvl+1)	**[-]**	-29.82	-21.26	-45.62	-21.59	-20.23	-7.846
		(5.68)***	(3.23)***	(5.18)***	(3.14)***	(3.30)***	(1.33)
... × (sd(U othr attr)-mean sd)	**[+]**	-	-	96.09	17.12	94.09	78.25
				(1.83)*	(0.80)	(4.95)***	(4.62)***
... × (sd(U this attr)-mean sd)	**[-]**	-	-	102.2	-10.95	-43.3	-116.5
				(1.99)**	(0.61)	(3.78)***	(12.04)***
... × (sasubrsk-msasubrsk)		-	-40.67	-	-41.03	-	-
			(7.48)***		(7.07)***		
... × (cosubrsk-mcosubrsk)		-	30.53	-	30.45	-	-
			(3.84)***		(3.78)***		
... × (benefits never)		-	217.4	-	215.8	-	-
			(6.63)***		(6.42)***		
... × (min overest latency)		-	8.219	-	8.415	-	-
			(13.65)***		(12.54)***		
Observations		21111	21111	21111	21111	21111	21111
Log L		-10992.674	-10326.046	-10976.589	-10316.015	-10915.004	-10771.54
Iterations				30	30	30	30
Marginal WTP for decr. in *log(pdvi+1)*							
5%		$ 5.73		$ 3.21		$-3.99	$0.08
50%		8.6		16.47		13.35	2.97
95%		11.72		27.35		13.02	6.61
Marginal WTP for decr. in *log(pdvl+1)*							
5%		$ 6.73		$ 8.59		$ 4.62	$0.73
50%		9.48		10.84		6.58	4.18
95%		12.81		15.24		28.68	8.03

Absolute value of z statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Lost life-years confer negative utility, and more so the greater the individual's subjective risk of the illness or injury in question. Again, the higher the average subjective risk rating for the other major illnesses in the survey, the lesser the disutility from lost life-years due to the cause in question. If the individual expects never to benefit, disutility is far less, and the greater the extent to which latency is overestimated, the lesser the disutility from lost life-years from a given cause. All of the systematic heterogeneity we identify is thus plausible.

Model SD3 uses a specification analogous to Model SD1 to calculate both the (latent) utility differences based on all other attributes and the utility differences based only on the attribute in question. We use 30 iterations of the model, by which point the permutation vector essentially disappears. In this specification, however, the systematic variation in estimated marginal utilities with respect to the two factors which the theory suggests should drive the relative attention to different attributes is not consistent with our theory. Only three of the key eight parameters bear the expected sign, and of these, only one is statistically significantly different from zero. Three are statistically significant but bear the incorrect sign.

We suspect that these disconcerting results stem from our assumption of homogeneous preferences in Model SD3. If everyone shares the same set of four utility parameters, then there will no variation across individuals in the utility derived from any one level of a given attribute or from any one set of levels for the rest of the attributes. So we allow for heterogeneous preferences. We use an analogy to Model SD2 in our iterative estimation process, with the results displayed as Model SD4. Again the results are disappointing, based on what we expect from our theory. Only two of the eight key parameters (those on the income term) are statistically significant and both of these bear the "wrong" sign.

However, we explore one additional specification, Model SD5. In this case, we suppose that the researcher adheres to the simple four-parameter specification used in Model SD1. Specifications such as this one—linear and additively separable in the list of attributes—are widely used in the choice literature. However, we use the richer, heterogeneous-preferences specification of Model SD2 to calculate the special variables used to control for the similarity of alternatives based on all utility implied by other attributes, as well as for the dissimilarity of alternatives based only on the utility implied by the attribute currently in question. In this case, the predictions of our theoretical model jump into sharp relief. All four basic utility parameters in the model bear the anticipated signs, seven of the eight key attention-related shift parameters also bear the anticipated signs, and six of these are statistically significant. The pairs of shifters for the sick-years and lost life-years variables are all strongly statistically significant (at the one percent level).

Model SD5 appears to tell a cautionary tale. Parameter estimates from the simple four-parameter homogeneous-preferences ("naive") specification in Model SD1 have the potential to be substantially biased by neglect of our two types of attention factors. By how much? The standard deviations which we use to measure the attention factors are normalized on their mean values in the sample, so that each baseline coefficient in Model SD5 corresponds to a case where the similarity of alternatives based on other attributes, and their dissimilarity based on the current attribute, are equal to their sample mean values across all respondents. Presumably, the baseline parameters at the means of the sample are approximately what we measure in Model SD1. Our results suggest that the combination of other-attribute dissimilarity and own-attribute dissimilarity influence respondent attention and result in a tendency for the researcher to overestimate the marginal utility derived from net income, overestimate the marginal disutility of sick-years and lost-life-years, and perhaps to underestimate the marginal disutility of recovered-years, on average.

The main insight from Model SD5 is that our usual models assume that the mix of attribute levels across alternatives should have *no* systematic effect on marginal utilities. The evidence in Model SD5 suggests that our theoretical insights may be supported in these

data, although it may be necessary to ensure a reasonable degree if individual-specific variation in preferences to capture the extent to which a set of alternatives may be judged (by each different individual) to be more or less similar based on their various attributes. Our example also suggests that, under a model with sufficient heterogeneity in preferences, these biases may be avoided, but we find evidence of their presence in a too-simple homogeneous-preferences specification.

The bottom panel of Table 2 summarizes the sizes of the distortions which may be due to respondents' attention to attributes having potentially been steered by the design of the choice set. Adverse health states are "bads," so the marginal utility of an additional "log discounted year" in any adverse health state can be assumed to be negative (i.e. it is a disutility). We convert each of our estimated marginal utilities to a marginal willingness to pay by dividing through by the marginal utility of net income. At the bottom of Table 2, based on 21111 draws from the assumed joint normal distribution of the maximum likelihood estimated parameters, we report the 5th, 50th, and 95th percentiles of the distribution of the ratio of these marginal utilities. Since the results for sick-years and lost life-years are the most robust, we report on the distributions of marginal WTP for a one-unit change in $\log\left(pdvX_i^j + 1 \right)$ for discounted sick-years and discounted lost life-years. If we assume that negative WTP for a worse health state is symmetric with the positive WTP for similar improvement in health status, then the simple four-parameter homogeneous preferences Model SD1 yields a median estimate of $8.6 for this change in discounted sick-years, whereas Model SD5 suggests that this marginal WTP is more like $13.35. For the discounted life-years variable, Model SD1 suggests a median estimate of $9.48, whereas Model SD5 implies a marginal WTP of only $4.62. Differences of this size are likely to be relevant to policy-making.

4.4.2 Alternative models using $lead\left(x'_{-ki}\beta_{-k} \right)$ and $lead\left(x_{ki}\beta_k \right)$:

The last column of Table 2 displays results for a model analogous to Model SD5 where the apparent marginal utility from each attribute is allowed to shift systematically with (1) the extent to which there is a "clear winner" among the three alternatives based on the other attributes; and (2) the extent to which there is a "clear winner" among the three alternatives based on the contribution made to utility by this attribute. This final column records that all coefficients in Model Lead5 have the anticipated signs, and furthermore, that the four key shift coefficients for sick-years and lost life-years are strongly statistically significantly different from zero at the one percent level. In these specifications, however, neither of the attention-related shifters for the income term is statistically significant (although each bears the anticipated sign).

In terms of the extent of the bias in the four parameters of the simple specification due to failure to account for deviations from the average across the three alternatives in the two measures of utility similarity and dissimilarity (based on other attributes and on the attribute in question), model Lead5 also suggests that the marginal utility of income may be overestimated in the naïve homogeneous-preferences model. The disutility from sick-years and lost life-years may be overestimated rather substantially, and the disutility from recovered years may be underestimated. Implications for the two main marginal WTP estimates in the model are again described in the bottom panel of the table.[15]

[15] Negative estimates of WTP are merely an artefact of the functional form. Respondents were given no opportunity to express negative WTP for any of the health-risk reduction programs in the survey, so we tend to interpret point estimates which imply negative WTP for improvements as merely zero

4.4.3 Other models

In this research, we have also explored a specification using $skew\left(x'_{-ki}\beta_{-k}\right)$ and $skew\left(x_{ki}\beta_{k}\right)$ and we have investigated models which use both standard deviation and skewness measures at the same time. Finally, we have examined specifications in terms of entropy measures. These other specifications appear to be less appropriate for the data used in our example, although they may be useful in other empirical applications.

It is reasonable to ask whether the distortions due to the mix of own- and other-attribute levels could be avoided by resorting to mixed logit models or models where the logit parameters are all random (and either uncorrelated or correlated). Recognition of some type of heterogeneity is usually preferable to ignoring heterogeneity altogether. We have estimated the specification reported as Model SD5 in Table 2, but with the four baseline coefficients in the naive model implemented as normally distributed random parameters. All four standard deviations for these random parameters are strongly statistically significantly different from zero. The expected values of the random parameters remain the same order of magnitude as the fixed coefficients in Model SD5, although the α_2 and α_3 coefficients remain significant only at the 10 percent level. However, our findings for the eight key shifters (four each for other-attribute dissimilarity and own-attribute dissimilarly) remain the same. All coefficients are strongly significant and bear the theoretically expected signs (except for the two shifters for the recovered-years term, which are also statistically insignificant in Model SD5).

4.4.4 Illustration: effects on marginal WTP estimates

Our finding that own-attribute and other-attribute dissimilarity measures can have a strongly statistically significant effect upon the estimated marginal utilities in discrete choice models is notable. Standard random utility models assume these effects are zero, so that the apparent marginal utilities from a discrete choice model are interpreted as being identical to the true underlying marginal utilities. We argue here, however, that these apparent marginal utilities are likely to represent a combination of true marginal utilities and attention to each attribute. Thus, it is also important to evaluate the potential effect of these drivers of systematic differences in attention on the resulting estimates of willingness to pay.

In our illustrative example, we know from other work with these data that a simple linear-in-logs specification tends to be too simple, so we will avoid calculating and advertising estimates of willingness to pay for health risk reductions programs as a package. However, we note that many researchers using discrete choice models tend to begin with simple linear models such as the one we entertain for our example. We thus calculate the likely scope of the systematic effects of our new dissimilarity variables on the implied estimates of marginal WTP for different attributes by examining their effects on the marginal rate of substitution between selected attributes and money.

Table 3 begins by reiterating, in its first row, the baseline point estimates (for Model SD5 in Table 2) of marginal utility associated with each (constructed) attribute in our model. These are the "levelized" marginal utilities which apply in the counterfactual instance where all four other-attribute utility-difference variables (and all four own-attribute utility-difference variables) are set equal to their sample means. These can be compared to our uncorrected naive model SD1, where homogeneous preferences are assumed. Whereas the

WTP. This is analogous to the intuition behind standard maximum likelihood Tobit models. While the latent variable may be negative, its observable manifestation is censored at zero.

marginal (dis)utility of a present discounted adverse health-state year ranged roughly between -24 and -29 for each of the three health states in the naive model, the levelized marginal utilities from model SD5 are considerably different.

The most important insights, however, may be gleaned from the rest of Table 3. Holding the own-attribute dissimilarity measure at its sample mean value, we first calculate and display the range of fitted values across selected percentiles of the sampling distribution of calculated other-attribute dissimilarity levels (based on the heterogeneous-preferences specification of Model SD5). The marginal utility of income parameter can be seen to differ by a factor of ten between the 5th and 95th percentiles of the distribution of other-attribute dissimilarity. Since this marginal utility enters into the denominator of WTP calculations, this implies a corresponding possible ten-fold difference in WTP estimates! If we instead hold constant the sample mean the other-attribute dissimilarity measure, we find that own-attribute utility differences can contribute to a 2.5-fold difference in the marginal utility of income parameter.

With respect to the other marginal utilities in the basic model, there is further evidence of heterogeneity stemming from differences in the mix of attribute levels used across the alternatives in the choice set. Since the means and medians differ, there is certainly some skewness in the estimated heterogeneous values. The presence of cross-overs in the signs of the estimated marginal utilities within the range of the data suggest that non-linear models may be desirable, since they could constrain the disutilities of adverse health states to be strictly negative. For this illustration, however, we seek merely to demonstrate that our two dissimilarity variables are potentially important shifters of the fitted values of the apparent marginal utility parameters in such a model.

4.5 Caveats

An anonymous referee has suggested that with enough data, one might contemplate defining an explicit objective function for attention. This objective function would be accompanied by an explicit cognitive budget for each individual that constrains the allocation of this limited amount of attention across different dimensions of a choice set, including attributes and alternatives as well as other features. The theory portion of this paper has assumed such an underlying optimization process, without making it explicit. The outcome of such a constrained optimization can be expected to yield the usual optimality condition that the expected marginal benefits of additional attention to a particular feature of a choice set should be set equal to the marginal costs of additional attention. Utility can be improved by more attention to some feature of the choice set as long as the expected marginal benefit from additional attention exceeds the marginal cost of additional attention.

In this paper, out of necessity, we have assumed that the marginal cost of attention to each different attribute is essentially identical across all choices and all respondents, although this is a very strong assumption. At least every choice set has the same number of alternatives and attributes, displayed in the same type of choice table, and the order of the attributes is held constant across choice sets. Attention to any given attribute should increase as long as marginal net benefits from attention are positive. If marginal costs of attention are the same everywhere, then optimal attention to any given attribute will be greater when expected marginal benefits are greater, assuming that they are large enough to exceed marginal costs. We must leave such an explicit attention-optimizing model for future research where the marginal costs of incremental attention to attributes (and other choice set features) can be more explicitly quantified.

The same anonymous referee noted that many strategies require less attention than a fully compensatory strategy. This referee asked whether it might be possible to implement a

Table 3: Sizes of the effects of dissimilarity variables on estimated marginal utilities (Model SD5)

	Denominator of WTP ↓	In numerator of WTP		
Dissimilarity variables normalized so that sample mean = 0	Income term (β_1)	Sick-years term (β_2)	Recovered-years term (β_3)	Lost life-years term (β_4)
MU at "mean" dissimilarity =	1.514	-7.124	-33.92	-20.23
Effects of *other*-attribute utility dissimilarity (percentiles):				
5th	2.24	-38.04	-49.38	-47.83
25th	2.01	-28.85	-44.52	-38.91
50th	1.68	-16.37	-37.64	-27.48
75th	1.19	6.03[a]	-27.63	-8.16
95th	0.21	50.81[a]	-4.99	30.20[a]
Effects of *own*-attribute utility dissimilarity (percentiles):				
5th	1.01	20.08[a]	-35.90	-7.53
25th	1.12	12.77[a]	-35.90	-10.76
50th	1.33	1.08[a]	-35.90	-15.94
75th	1.70	-18.80	-33.47	-25.07
95th	2.67	-62.93	-25.82	-48.23

[a] Unexpected signs on some of these fitted marginal utilities result from estimation without constraints. In a non-linear adaptation of this model, it would be possible to estimate the negative of the logarithm of each marginal utility, and to allow this log-transformed parameter to shift systematically with the two types of dissimilarity measures. This would constrain the fitted marginal utility to remain strictly negative.

different framework, where attention was applied to tradeoffs between attributes, rather than to the attributes themselves. In a linear and additively separable model with four basic attributes, there are six possible ratios of marginal utilities (marginal rates of substitution) that respondents might consider. Unfortunately, it is beyond the scope of this paper to develop and implement an alternative specification where six attention parameters are associated with these six ratios of the four basic marginal utility parameters, so we leave this option for subsequent research as well. Such a specification would of course be interesting, since this referee notes that it might permit us to discriminate between fully compensatory strategies, conjunctive strategies, disjunctive strategies or even simply an adding up of attributes that exceed a threshold. This alternative approach might also be valuable if individuals actually have lexicographic preferences (e.g. always just choose the cheapest option—although in our example, this would be the status quo in every choice set). But consider a case of forced choice, where every alternative comes at a cost. If the individual chooses solely on the basis of cost, and costs are identical, he or she may pick randomly

among alternatives with these equal costs. Our model would not perform well if everyone chose in this manner.

We resort to our "alternating estimator" because of the fundamental challenge of identification when the same set of utility parameters shows up in so many places in the objective function for FIML estimation. It has been suggested that we might consider breaking the link between the preferences used in the utility function and the preferences used in the attention component of the model by adopting the equal weights proposed by Dawes (1979). A system of equal weights would fix the "utility" coefficients in the calculation of the attention variables arbitrarily at unity. We feel that our alternating estimator is preferable, however, since it fixes these coefficients at their estimated values from the previous iteration of the richer heterogeneous-preferences specification for the "true" underlying utility function (which we assume would be unobserved by the typical practitioner in search of a simple linear and additively separable model for the representative consumer).

We note that it is tempting to consider using time-on-task durations for each choice as a direct proxy for attention to that choice. In a laboratory setting, this might be viable. For our internet-based field survey, it would be a risky strategy. A longer duration on a choice task can just as easily mean that the respondent was distracted from the task for a few seconds up to a few hours. In a laboratory setting, one might also use eye-tracking software to measure how long the respondent's gaze dwells upon an area of the screen occupied by each attribute. With this more-direct information about attention, it would be much easier to build explicit constraints for the attention-related terms in our model. With the current data, however, it is very difficult to come up with additional constraints to aid in the identification of the attention and utility-related parameters.[16]

In our empirical application, it is much more convenient to use the first variant of equation (12) for the propensity to attend to different attributes. Had it been equally tractable to employ the third variant, it would be straightforward to assess whether the attention factor could be equal across all attributes for any given respondent. This would suggest a model equivalent to one wherein the scale factor for the utility-difference function merely differs across people. This type of test is left for future research where the third variant in equation (12) can be implemented.

Subjects in our study each have the opportunity to make five different choices. In this case, it may not be the standard deviation across the current choice set in utility contributions for a particular attribute which determines attention to that attribute. Instead, it may be the standard deviation in utility contributions across both the current and all previous choice sets that determine the attention devoted to an attribute. We do not pursue this possibility here.

To allow the baseline marginal utility parameters to be comparable across our various specifications, we first normalize our dissimilarity variables on their mean values across all respondents. This permits us to consider the case where all dissimilarity variables might match the sample-wide mean as equivalent to the case where the shift variables we actually use are all simultaneously zero. We have not addressed the possibility that one might normalize on the within-individual means, but to allow these mean dissimilarity measures to differ across individuals. An a priori sense of the promise of such a strategy is harder to come by, since the relevant quantities are factors in interaction terms, rather than basic variables in the model.

[16] Chabris et al. (2009) consider the allocation of time across choice tasks according to the "value gap" between the two options in each choice. However, they do not consider the allocation of attention among attributes, or among alternatives within a choice task.

One might argue that our use of the richer heterogeneous-preferences model to generate our eight fitted dissimilarity measures is merely an alternative strategy for bringing respondent heterogeneity into the naive homogeneous preferences model. This criticism may be supported by the fact that the same dissimilarity measures make no real difference when they are added to the heterogeneous-preferences model. But this does not take away from the intriguing finding that respondent heterogeneity—exclusively via its influence on the two types of theoretically motivated measures of alternative similarity—contributes so very much to explaining differences in apparent marginal utilities in the naive model (which is where many practical conjoint choice analyses begin and end). [17]

5 Conclusions and Potential Implications

In conventional random utility choice models, researchers usually assume complete and costless information. However, subjects' cognitive resources are typically scarce. Individuals presumably must compare the expected marginal benefits and marginal costs of attention to different dimensions of a choice task, and optimize their allocation of attention. In this paper, we focus on the individual's allocation of his or her attention across the different attributes which can be used to describe each alternative in a choice set. Inattention to differences in the levels of a particular attribute may masquerade empirically as a lower marginal utility associated with that attribute. Marginal utilities from choice models are the key ingredients in the calculation of willingness-to-pay in many applications. Distortions in these marginal utilities can lead to distortions in the sorts of willingness-to-pay estimates which are critical to an understanding of demands for the goods in question.

Our illustrative empirical example represents a first partial attempt to implement an attention-corrected choice model with a sample of "field" data from a conjoint choice experiment in a large stated preference survey. When we use a four-parameter homogeneous-preferences model to build the two dissimilarity measures associated with each attribute, and use these two measures to shift each marginal utility in what is otherwise the same four-parameter homogeneous model, we find no evidence of the effects predicted by our theory. We then generalize our model to make each of our four marginal utilities a systematically varying parameter, allowing for heterogeneity in preferences. If these heterogeneous preferences are used to build the two dissimilarity measures associated with each attribute, and these measures are the used to shift each of the four marginal utilities in the same heterogeneous-preferences model, they likewise fail to produce the effect predicted by our theory.

However, we subsequently assume heterogeneous preferences in the process of constructing the dissimilarity measures, so that our pairs of dissimilarity measures associated with each attribute differ across individuals because their preferences are different. Using these heterogeneous dissimilarity measures as estimates of "latent" variables that have the capability to shift the four basic marginal utilities in a homogeneous-preferences model produces highly significant results fully consistent with our theory. Choice modellers do often explore homogeneous-preferences specifications, seeking to estimate preferences for a representative consumer. Our results certainly suggest that such "representative preferences" may be biased by heterogeneity in perceived dissimilarities.

Our theoretical and empirical explorations of criteria that may affect a respondent's optimal allocation of attention to attributes also have some implications for other regularities

[17] One can in principle impose sign restrictions by assuming, for example, a log-normal rather than normal distribution for each parameter (using the negative of the variable in estimation if a strictly negative coefficient is desired). For this application, however, models with such restrictions failed to converge.

which have been observed in different types of choice behaviour. Most of these issues are basically familiar to choice modellers. What our research introduces is an additional theoretical justification and attention-based rationale for why these patterns are observed. Our research also suggests the extent to which WTP estimates could be unintentionally biased by arbitrary decisions about the mix of attributes in a choice set. Our work is certainly not the last word on this subject, however, since the limited number of dimensions of variability across the choice sets in our empirical application do not permit us to test all of the possible predictions based on this type of approach to bounded rationality in choice situations.

SP too different from RP choice sets.—Choice set designs used for stated preference surveys may produce uneven attention to different attributes. This might be of little consequence if the corresponding real choice contexts were assured of being similar. However, if the conditions surrounding the choice are sufficiently different in the context wherein a choice prediction is desired—so that the marginal benefits and/or marginal costs of attention to attributes are different—a model calibrated under an implicit assumption of complete attention (when this is not so) may produce misleading forecasts of future choices. This suggests that when SP data are to be used to predict likely RP choices, it may be important to design those SP choice sets to feature the types of attribute mixes that will be encountered in the real world, to ensure that the distribution of attention across attributes in the SP case will be similar to that in the future RP case.

Consequentiality.—In purely hypothetical SP choice contexts, where stated choices may be viewed as inconsequential, the marginal benefits from attention to all attributes could be perceived to be very low (see Carson et al. 2003; 2004). In contrast, the marginal costs of attention to any additional attribute may be very similar to those in a real choice context. A lack of perceived consequentiality would thus be predicted to lead to a lower overall optimal level of attention being paid to the choice task. If attention to each attribute is reduced proportionally, we have argued that the only substantive effect may be observationally equivalent to an increase in the error dispersion, relative to the full attention case, with no resulting bias in the relative sizes of the estimated marginal utility parameters, and thus no distortion in any resulting estimates of the expected WTP. However, levels of attention to different attributes may not be scaled down uniformly across all attributes when less-than-complete attention is optimal. Our theory focuses on the marginal benefits part of the story, and suggests that the marginal benefits from attention to an additional attribute depend in a fairly complex fashion upon the pattern of attributes in the choice set and on the individual's marginal utilities from each attribute. Simply the design of a choice set can steer the subject's attention toward some attributes and away from others.

If the purpose of the SP choice task is to measure social preferences for a non-market public good, such as environmental quality, one may wish to simulate the preference parameter estimates that would emerge under the counterfactual case where everyone in the sample devotes their complete attention to every attribute in each choice set. These would simulate the "full information" case that would be highly desirable in the estimation of preferences in such a context.

Price listed last.—In SP experiments, the utility loss from making a wrong choice can be negligible, since the individual may believe that he or she will not have to live with the consequences of an "incorrect" choice—in particular, the knowledge that they have paid good money for something that turned out to be not exactly what they wanted or expected. If respondents do not fully take into account the fact that they would actually have to pay the cost of the preferred alternative, they may pay less attention than they should to any differences in costs (especially if these costs are listed at the bottom of the conjoint choice table). If order effects increase the relative cost of attention to the cost attribute, and attention is steered toward other attributes by listing them first, it may be unsurprising that

the propensity to attend to other attributes will be greater than the propensity to attend to cost. The marginal utility of income may be underestimated by more than the marginal utilities of the other attributes. The predicted result would be an upward bias in WTP, since the marginal utility of income forms the denominator in WTP calculations.

Cheap talk scripts.—In SP surveys, since the publication of Cummings and Taylor (1999), researchers have been encouraged to employ a so-called "cheap talk" script wherein subjects are specifically reminded to consider their budget constraint carefully before stating their preferred option. This section of a survey will typically draw special attention to the cost attribute, immediately prior to the choice task. In a conjoint choice context, this effort can be expected to increase attention to the cost attribute without treating the other attributes symmetrically. Our theory suggests that this can be expected to lead to a larger-than-otherwise estimated marginal utility of income and perhaps smaller-than-otherwise estimated marginal utilities for other attributes (if scarce attention is reallocated), which will tend to "bias," rather than "correct" the resulting WTP estimates. However, if a lack of consequentiality for the entire choice exercise has already produced lower attention to the cost attribute, the cheap talk effort may be warranted. However, our results strongly suggest that any attempt to direct the subject's attention specifically towards one attribute or another should be examined very carefully. We know that the cost attribute is frequently downplayed in RP contexts: restaurant menus list the price of the entree last, and advertisements encourage prospective customers to "contact the dealer for price information." Effort is often made to divert attention from other attributes as well, especially where they may convey negative marginal utilities: some less desirable attributes of goods for sale are listed in the fine print (e.g. pharmaceutical side effects),

SP attributes sometimes "too orthogonal."—To maximise estimation efficiency for marginal utilities associated with a whole range of attributes, the joint distribution of attributes in SP studies often has greater orthogonality or greater variance than might be present in the corresponding real-world choice context. As Jordan Louviere has pointed out, "Realism is not a design property" for a choice set. Our theoretical results suggest that the degree of orthogonality in attributes may have a systematic effect on the sizes of naively estimated marginal utilities. In the corresponding real choice context, subjects may face alternatives where the differences in many attributes across alternative may be much smaller than they had been in the SP estimating sample. This lesser difference changes the expected net benefits from considering the different attributes and changes the extent to which the individual is likely to take into account each of these attributes in the real-choice context. A choice model estimated on SP data would incorrectly predict choices under the different choice regimes in a subsequent RP setting.

More "ceteris paribus" than in real choices.—While attribute levels may be more different in some SP studies than they are in real life, in other cases the researcher's goal is merely to obtain a precise estimate of just one marginal utility. In this situation, the choice sets might consist of alternatives where all other attributes are held constant and only the attribute of interest is varied across alternatives. In some cases, the choice scenario may not even list other important attributes and will simply ask respondents to assume that all other features of the alternatives are identical. Our theory suggests that the greater the number of attributes held essentially constant across alternatives, the larger will be the apparent marginal utilities associated with the attributes which do vary.[18]

[18] For example, had the identical resumes with different names been sent to the same prospective employers in the Bertrand and Mullainathan (2004) study, one might expect that race, as implied by the different names, might have been found to have an exaggerated influence on choices compared to a choice context where the resumes differed in many other dimensions as well.

Marginal cost differences.—Our theory does not explicitly derive the factors which should determine the marginal cost of attention to an additional attribute. However, intuition suggests that the marginal costs of attention to different attributes may also differ. A variety of conditions could affect the marginal cost of attention to an incremental attributes. One is the accessibility of the information about each attribute (e.g. its position in the order of attributes in a conjoint choice scenario). For a decision-maker faced by different levels of distraction or time pressure in making a choice, the fact that the marginal cost of attention is likely increasing in the number of attributes in a choice set will also be relevant. Some attributes, such as risk for example, may be more difficult to understand for some types of subjects. This suggests that there will likely be differences in the marginal propensity to attend to each attribute whenever cognitive constraints are binding. Differences in attention can lead to biases in estimated marginal utilities and thereby to distortions in estimated WTP.

We have derived, from optimizing behaviour, results that seem to match closely with casual empiricism about how people make choices. Individuals are motivated to pay attention to additional attributes in a choice exercise to the extent that this behaviour will reduce their expected lost utility from making an incorrect choice. They pay more attention to any given attribute if the alternatives look more similar in terms of utility based on the other attributes under consideration. They also pay more attention to an attribute if the utility derived from that attribute differs greatly across alternatives. These are simple insights. In our empirical adaptation of this theory, we encounter some difficulty in estimation of an appropriate specification using full-information maximum likelihood methods. This is because the same utility parameters appear in so many places in the log-likelihood. Nevertheless, we have implemented the estimation in an alternating sequence of steps that appears to lead to stable converged parameter estimates. We demonstrate that the apparent marginal utilities from different attributes can vary dramatically with the mix of attribute levels presented across all alternatives, and thus so can the implied WTP. This is more evidence that, by manipulating the mix of attribute levels in a choice set, it may be possible to "steer" respondent attention (inadvertently or strategically) to either exaggerate or downplay apparent marginal utilities and hence the resulting average WTP (benefits estimates). Our findings may therefore have important implications for how researchers approach the problem of experimental design in specifying choice sets in SP research. They may also explain problems in using RP data from one type of choice context to infer likely behaviour in another context where the patterns of attributes across alternatives are too different.

6 Acknowledgements

Our recognition of the need for this paper was crystallized by discussions in the session entitled "Dissecting the Random Component" at the Fifth Triennial Invitational Choice Symposium hosted by UC Berkeley at Asilomar in Pacific Grove, CA, in June 2001. Valuable feedback was obtained from presentation of the theory section in the session entitled "Recent Progress on Endogeneity in Choice Modeling" at the Sixth Triennial Invitational Choice Symposium in Estes Park, CO in June 2004. Initial empirical results supporting the model were presented in the session entitled "Behavioral Frontiers in Choice Models" at the Seventh Triennial Invitational Choice Symposium at the Wharton School in Philadelphia, PA, in June 2007. We are grateful to all of the participants at these sessions who have influenced our thinking on this topic, as well as to members of the Triangle Resource and Environmental Economics seminar sponsored by NC State, RTI, and Duke University, and the 10th Occasional Workshop on Environmental and Resource Economics

at UCSB. Jason Lindo has provided some helpful editorial comments. This research was supported in part by the National Science Foundation (SES-0551009) and by the Raymond F. Mikesell Foundation at the University of Oregon.

7 Appendices

7.1 Alternating Estimation Algorithm

Step 0: Estimate the model without any attention corrections. Save these temporary estimates of the *k=1,...,K* marginal utility parameters. These might be scalars, $(\hat{\beta}_1^0,...,\hat{\beta}_K^0)'$, or systematically varying parameters, $(\hat{\beta}_1^{0'}Z_{1i},...,\hat{\beta}_K^{0'}Z_{Ki})'$, each depending on a sub-vector of parameters $\hat{\beta}_k^0$ and a vector Z_{ki} of individual characteristics.

Step 1: Based on these initial estimates of the four marginal utilities, construct the contribution to net indirect utility associated with each attribute, relative to that for the numeraire alternative, *J*. This may be a single scalar marginal utility times its associated attribute level, $x_{ki}^j\hat{\beta}_k$,or it may be a systematically varying marginal utility times the associated attribute level, $x_{ki}^j(\hat{\beta}_k'Z_{ki})$, for *k=1,...,K*. Sum these contributions across all attributes to calculate $x_i^{j'}\hat{\beta}$, the net indirect utility index associated with each alternative, *j=1,...,J-1*. For the numeraire alternative, this net utility will be zero.

a.) For each attribute, subtract from total systematic utility the contribution made by just that attribute to leave the model's prediction about net indirect utility based only on the other attributes in the model, $x_{-ki}^j\hat{\beta}_{-k}$ (or $x_{-ki}^j(\hat{\beta}_{-k}'Z_{-ki})$ in the systematically varying parameter case). Construct a measure of the dissimilarity of the alternatives on the basis of these other attributes, $dissim(x_{-ki}\hat{\beta}_{-k}^0)$, or $dissim(x_{-ki}(\hat{\beta}_{-k}^{0'}Z_{-ki}))$. We have suggested several candidates: the size of the lead, in utility units, the standard deviation across alternatives in these other-attribute utility levels, and the skewness in these measures across alternatives. Adjust the location of these measures by using their deviations from the overall sample mean values (or any other target value to be simulated by zeroing out this dissimilarity measure): $d\_dissim(x_{-ki}\hat{\beta}_{-k}^0) = dissim(x_{-ki}\hat{\beta}_{-k}^0) - mean\_dissim(x_{-ki}\hat{\beta}_{-k}^0)$.

b.) For each attribute, construct a measure of the dissimilarity of the three alternatives on the basis of just this attribute: $dissim(x_{ki}\hat{\beta}_k^0)$ or $dissim(x_{ki}(\hat{\beta}_k^{0'}Z_{ki}))$. Again, possible candidates include the lead of the highest utility contribution due to this attribute, over the second-highest across alternatives, or the standard deviation, or the skewness in these utility-contributions across alternatives. Again, adjust these measures by using their deviations from the overall sample mean values (or some other target value to be simulated when the deviations are all zero), to yield $d\_dissim(x_{ki}\hat{\beta}_k^0)$.

Step 2: Re-estimate the model, but now allow the marginal utility from each attribute (or the intercept of the marginal utility expression, if it is modelled as a systematically varying parameter) to vary systematically with the calculated dissimilarity of the alternatives in this choice set based on net utility from other attributes, as well as the dissimilarity of the alternatives based on net utility only from this attribute. Each "observed" marginal utility parameter is now modelled as also varying systematically with $dissim(x_{-ki}\hat{\beta}_{-k}^0)$ and

$dissim(x_{ki}\hat{\beta}_k^0)$. In this second iteration, the new vector of "true (corrected)" underlying marginal utility parameters for each attribute, $\hat{\beta}_k^1$, is supplemented by the estimated coefficients on each of these two dissimilarity terms, yielding $(\hat{\beta}_k^1, \hat{\alpha}_k^1, \hat{\theta}_k^1)$ for *k=1,...,K*. If the marginal utilities in the model are scalars, this generalization will triple the number of estimated parameters. If the marginal utilities are systematic varying parameters, the number of estimated parameters will increase by *2K*.

Step 3: Net out the estimated biases in systematic utility due to $d\_dissim(x_{-ki}\hat{\beta}_{-k}^0)$ and $d\_dissim(x_{ki}\hat{\beta}_k^0)$ by setting these *2K* different constructed variables to zero. This simulates the case where, for all attributes, the dissimilarity of alternatives based on all other attributes, and based on each specific attribute, is the same for all attributes in all choice sets. We then interpret the other utility parameters in the model as the "true" utility parameters (corrected for attention biases created (unintentionally?) by the mix of attributes designed into the choice set).

Step 4: Repeat Step 1, now using these updated estimates of the basic utility parameters, $(\hat{\beta}_1^1,...,\hat{\beta}_K^1)'$, or systematically varying parameters, $(\hat{\beta}_1^{1'}Z_{1i},...,\hat{\beta}_K^{1'}Z_{Ki})'$, as the "true" utility parameters to construct updated measures of dissimilarity, $d\_dissim(x_{-ki}\hat{\beta}_{-k}^1)$ and $d\_dissim(x_{ki}\hat{\beta}_k^1)$. Continue to iterate through Step 1 through 3 until the length of the step-to-step permutation in the parameter vector becomes arbitrarily small.

7.2 Corrected Variance-Covariance Matrix

We have also estimated Models SD1 through SD5 as conditional logit-type specifications without fixed effects, with results as shown in Table A. A fixed effects specification is less crucial in this context because the attributes of the alternatives in our choice sets were randomized, subject only to exclusions for implausibility. Results of a similar flavour emerge, with the analogue to Model SD5 again providing evidence of the types of attention-diverting effects suggested by our theory. These non-fixed-effects models also allow us to address the problem that the standard errors at the last iteration of the steps in the estimation algorithm do not reflect the fact that the variables used for the eight dissimilarity terms are calculated based on the last round of point estimates from the heterogeneous specification (which also involves fitted dissimilarity variables from the most recent round of estimates).

Ideally, one would estimate all parameters of the model simultaneously by full information maximum likelihood. However, since the basic utility parameters appear in so many different places in these models, we are not surprised to find that such likelihood function is very difficult to optimize by standard methods. Using the alternating algorithm, convergence seems to be straightforward and unambiguous. When the converged point estimates from the alternating algorithm are inserted into the full likelihood function for the same problem and numerical derivatives are calculated for the full set of parameters, there is some shrinkage of the asymptotic *t*-test statistics on most parameters, but everything that was statistically significant at better than the 10 percent level at the end of the alternating algorithm remains significant in terms of the full log-likelihood function. However, we note one markedly larger t-test statistic for the very last parameter in the model. The standard step-sizes for numeric derivatives may be inappropriate for this parameter. This particular test statistic needs yet to be understood.

Table A: "Standard Deviation" Variant: Uncorrected and Attention-corrected **Non-Fixed-Effects** Conditional Logit Models for Health-Risk Reduction Programs (with homogeneous and heterogeneous preferences)

		(SD1) iterative	(SD2) iterative	(SD3) iterative	(SD4) iterative	(SD5) iterative	(One-step eff.) FIML
	Exp. sign	Homogeneous preferences	Heterogeneous preference	Attention homogenous-homogeneous	Attention heterogeneous-heterogeneous	Attention heterogeneous-homogeneous	Attention heterogeneous-homogeneous
Income term (β_1)							
(income term)	**[+]**	3.364 (8.28)***	3.72 (6.66)***	4.769 (0.38)	3.222 (3.37)***	2.475 (2.78)***	2.475 (2.22)**
...× (sd(U othr attr)-mean sd)	**[-]**	-	-	58.84 (2.05)**	3.646 (3.16)***	-2.227 (2.30)**	-2.227 (-1.94)*
... × (sd(U this attr)-mean sd)	**[+]**	-	-	470.6 (1.75)*	-3.995 (1.91)*	2.256 (1.64)	2.256 (1.03)
…× female		-	3.253 (5.00)***	-	4.952 (4.93)***	-	-
Sick-years term (β_2)							
$\Delta\Pi_i^{AS}$ log(pdvi+1)	**[-]**	-28.87 (4.76)***	-17.8 (2.37)**	-14.3 (0.78)	-17.77 (1.34)	-18.57 (1.59)	-18.57 (-1.7)*
... × (sd(U othr attr)-mean sd)	**[+]**	-		-29.67 (0.65)	-22.4 (1.00)	109.9 (5.83)***	109.9 (4.46)**
... × (sd(U this attr)-mean sd)	**[-]**	-		-71.4 (0.37)	11.54 (0.45)	-121.1 (6.47)***	-121.1 (-6.89)**
... × (sasubrsk-msasubrsk)		-	-22.49 (3.83)***	-	-26.18 (4.27)***	-	-
... × (cosubrsk-mcosubrsk)		-	29.68 (3.47)***	-	33.21 (3.80)***	-	-
... × (benefits never)		-	126.4 (3.86)***	-	122.3 (3.65)***	-	-
... × (min overest latency)		-	7.614 (11.85)***	-	8.456 (11.36)***	-	-
Recovered-years term (β_3)							
$\Delta\Pi_i^{AS}$ log(pdvr+1)	**[-]**	-24.29 (2.53)**	-40.35 (3.92)***	-76.69 (2.79)***	-30.9 (1.23)	-34.08 (1.78)*	-34.08 (-1.57)

... × (sd(U othr attr)-mean sd)	**[+]**	-	-	-19.88	-14.13	23.73	23.73
				(0.21)	(0.48)	(1.04)	-0.69
... × (sd(U this attr)-mean sd)	**[-]**	-	-	127.6	-38.32	-29.48	-29.48
				(2.15)**	(0.26)	(0.31)	(-0.611)
... × (age-mage)		-	-1.614	-	-1.335	-	-
			(2.40)**		(1.37)		
Lost life-years term (β_4)							
$\Delta\Pi_i^{AS}$ log(pdvl+1)	**[-]**	-30.73	-23.03	-57.94	-16.33	-43.93	-43.93
		(5.88)***	(3.50)***	(3.88)***	(1.43)	(4.43)***	(-4.73)**
... × (sd(U othr attr)-mean sd)	**[+]**	-	-	84.86	-2.609	104.7	104.7
				(1.54)	(0.10)	(4.74)***	(3.8)***
... × (sd(U this attr)-mean sd)	**[-]**	-	-	64.72	-23.58	-36.28	-36.28
				(1.55)	(1.20)	(2.92)***	(-36.8)*** [a]
... × (sasubrsk-msasubrsk)		-	-40.74	-	-40.28	-	
			(7.44)***		(6.97)***		
... × (cosubrsk-mcosubrsk)		-	33.15	-	32.62	-	-
			(4.19)***		(4.06)***		
... × (benefits never)		-	204.2	-	209.9	-	-
			(6.33)***		(6.35)***		
... × (min overest latency)		-	7.869	-	8.019	-	-
			(13.11)***		(12.31)***		
Observations		21111	21111	21111	21111	21111	21111
Log L		-7682.953	-7050.867	-7670.404	-7042.261	-7603.081	-14645.34
Iterations				40	40	40	1
Marginal WTP for incr. in *log(pdvi+1)*							
5%		-11.37		-10.32		-18.1	
50%		-8.58		-3.31		-17.72	
95%		-5.92		9.06		-9.9	
Marginal WTP for incr. in *log(pdvl+1)*							
5%		-11.97		-34.65		-41.43	
50%		-9.11		-.93		-7.49	
95%		-6.64		32.46		.34	

Absolute value of z statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

[a] The reason for this unexpectedly small standard error (large t-test statistic) is unclear. These standard errors are calculated by substituting the converged values of the parameters from the sequential method into the full maximum likelihood function, followed by calculation of numeric derivatives at this optimum.

References

Bertrand, M. and S. Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, *American Economic Review* 94 (4), 991-1013.

Bettman, J. R., E. J. Johnson, M. F. Luce and J. W. Payne. 1993. Correlation, conflict, and choice, *Journal of Experimental Psychology: Learning, Memory and Cognition* 19 (4), 931-951.

Bettman, J. R., M. F. Luce and J. W. Payne. 1998. Constructive consumer choice processes, *Journal of Consumer Research* 25 (3) 187-217.

Cameron, T. A. and J. R. DeShazo. 2009. Demand for Health Risk Reductions, manuscript, Department of Economics, University of Oregon.

Carson, R., T. Groves, J. List and M. Machina. 2004. Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions Underlying Stated Preferences, Discussion Paper, Department of Economics, University of California, San Diego.

Carson, R., T. Groves and M. Machina. 2003. Incentive and Informational Properties of Preference Questions, Discussion Paper, Department of Economics, University of California, San Diego.

Conlisk, J. 1996. Why bounded rationality? *Journal of Economic Literature*, 34 (2), 669-700.

Chabris, C. F., C. L.Morris, D. Taubinsky, D. Laibson and J. P. Schuldt. 2009. The allocation of time in decision-making, *Journal of the European Economic Association* 7 (2-3), 628-637.

Chakravarti, A. and C. Janiszewski. 2003. The influence of macro-level motives on consideration set composition in novel purchase situations, *Journal of Consumer Research* 30 (2), 244-258.

Conlon, B., B. G. C. Dellaert and A. van Soest. 2001. Optimal Effort in Consumer Choice: Theory and Experimental Analysis for Binary Choice, discussion paper, Center for Economic Research (CentER), Faculty of Economics and Business Administration, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

Cummings, R. G. and L. O. Taylor. 1999. Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method, *American Economic Review* 89 (3), 649-665.

Dawes, R. M. 1979. The robust beauty of improper linear models in decision making, *American Psychologist*, 34, 571-582.

Dellaert, B. G. C., J. D. Brazell and J. J. Louviere. 1999. The effect of attribute variation on consumer choice consistency, *Marketing Letters* 10, 139-147.

DellaVigna, S. 2009. Psychology and economics: Evidence from the field, *Journal of Economic Literature* 47(2) 315-372.

de Palma, A., G. M. Myers and Y. Y. Papageorgiou. 1994. Rational choice under an imperfect ability to choose, *American Economic Review* 84, 419-440.

DeShazo, J.R. and G. Fermo. 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency, *Journal of Environmental Economics and Management*, 44 (1), 123-143.

Duquette, E., Cameron and DeShazo. 2009. Objective Choice Complexity versus Indirectly and Directly Measured Choice Difficulty in Discrete Choice Models, manuscript, Department of Economics, University of Oregon.

Fischer, G. W., J. Jia and M. F. Luce. 2000a. Attribute conflict and preference uncertainty: The RandMAU model, *Management Science*, 46 (5), 669-684.

Fischer, G. W., M. F. Luce and J. Jia. 2000b Attribute conflict and preference uncertainty: Effects on judgment time and error, *Management Science*, 46 (1), 88-103.

Gabaix, X. and D. Laibson. 2000. A boundedly rational decision algorithm, *American Economic Review* 90 (2), 433-438.

Gabaix, X., D. Laibson, G. Moloche and S. Weinberg. 2006. Costly information acquisition: Experimental analysis of a boundedly rational model, *American Economic Review* 96, 1043-1068.

Haab, T. C. and R. L. Hicks. 1999. Choice set considerations in models of recreation demand: History and current state of the art, *Marine Resource Economics* 14 (4), 255-270.

Heiner, R.A. 1985. Origin of predictable behaviour – further modelling and applications, *American Economic Review* 75 (2), 391-396.

Heiner, R. A. 1983. The origin of predictable behavior, *American Economic Review* 73 (4), 560-595.

Hensher, D. A. 2006a. How do respondents process stated choice experiments? Attribute consideration under varying information load, *Journal of Applied Econometrics* 21 (6), 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 (1), 7-44.

Hensher, D. A., J. M. Rose and T. Bertoia. 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 (2), 73-89.

Hensher, D. A., J. M. Rose and W. H. Greene. 2005. The implications on willingness to pay of respondents ignoring specific attributes, *Transportation* 32 (3), 203-222.

Jedidi, K. and R. Kohli. 2005. Probabilistic subset-conjunctive models for heterogeneous consumers, *Journal of Marketing Research* 42 (4), 483-494.

Johnson, E. J. 2008. Man, my brain is tired: Linking depletion and cognitive effort in choice, *Journal of Consumer Psychology*, 18 (1), 14-16.

Luce, M. F., J. Jia and G. W. Fischer. 2003. How much do you like it? Within-alternative conflict and subjective confidence in consumer judgments, *Journal of Consumer Research*, 30 (3), 464-472.

Malhotra, N. K. 1982. Information load and consumer decision making, *Journal of Consumer Research* 8 (4), 419-430.

March, J. G. 1978. Bounded rationality, ambiguity, and the engineering of choice, *Bell Journal of Economics* 9 (2), 587-607.

Mazzotta, M. J. and J. J. Opaluch. 1995. Decision making when choices are complex: a test of Heiner's hypothesis, *Land Economics* 71 (4), 500-515.

Paulssen, M. and R. P. Bagozzi. 2005. A self-regulatory model of consideration set formation, *Psychology & Marketing* 22 (10), 785-812.

Puckett, S.M. and D.A. Hensher. 2008. The role of attribute processing strategies in estimating the preferences of road freight stakeholders under variable road user charges, *Transportation Research Part E: Logistics and Transportation Review* 44 (3), 379-395.

Shugan, S. M. 1980. The cost of thinking, *Journal of Consumer Research* 7 (2), 99-111.

Simon, H. A. 1955. A behavioral model of rational choice, *Quarterly Journal of Economics* 69 (1), 99-118.

Swait, J. and W. Adamowicz. 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 (2), 141-167.

Swait, J. and W. Adamowicz. 2001b. The influence of task complexity on consumer choice: A latent class model of decision strategy switching, *Journal of Consumer Research*, 28 (1), 135-148.