Limits on Individual Choice

Sheshinski, Eytan

The Hebrew University of Jerusalem

March 2010

Limits on Individual Choice

by
Eytan Sheshinski*

May, 2010

Abstract

Individuals behave with choice probabilities defined by a multinomial logit (MNL) probability distribution over a finite number of alternatives which includes utilities as parameters. The salient feature of the model is that probabilities depend on the choice-set, or domain. Expanding the choice-set decreases the probabilities of alternatives included in the original set, providing positive probabilities to the added alternatives. The wider probability 'spread' causes some individuals to further deviate from their higher valued alternatives, while others find the added alternatives highly valuable. For a population with diverse preferences, there exists a subset of alternatives, called *the optimum choice-set*, which balances these considerations to maximize social welfare. The paper analyses the dependence of the optimum choice-set on a parameter which specifies the precision of individuals' choice (*'degree of rationality')*. It is proved that for high values of this parameter the optimum choice-set includes all alternatives, while for low values it is a singleton. Numerical examples demonstrate that for intermediate values, the size and possible nesting of the optimum choice-sets is complex. Governments have various means (defaults, tax/subsidy) to directly affect choice probabilities. This is modelled by 'probability weight' parameters. The paper analyses the structure of the optimum weights, focusing on the possible exclusion of alternatives. A binary example explores the level of 'type one' and 'type two' errors which justify the imposition of early eligibility for retirement benefits, common to social security systems. Finally, the effects of heterogeneous degrees of rationality among individuals are briefly discussed.

JEL Classification: C 35, D 03, D 81, H 80.
Key Words: Multinomial Logit, choice-set, default options, early eligibility constraint.

---

* Department of Economics and Center for the Study of Rationality, The Hebrew University of Jerusalem. E-mail: mseytan@mscc.huji.ac.il
This paper was presented at the conference celebration for Peter Diamond, MIT, May 1, 2010. I am grateful to Joram Mayshar for important comments.

1 Introduction

Early in modern psychology it has been observed (Luce and Suppes (1965)) that in choice experiments individuals do not select the same alternative in repetitions of identical situations. To explain these behavioral inconsistencies a probabilistic choice mechanism was introduced. Two alternative approaches have been offered as the foundation for probabilistic choice. One assumes that individuals have well-defined utilities but that instead of selecting the alternative with the highest utility, the decision maker is assumed to behave with choice probabilities defined by a probability distribution function over the alternatives that includes the utilities as parameters. This can be interpreted as the analysts' admission of lack of knowledge about individuals' decision processes. The selection of a probability distribution function is based on specific assumptions with respect to the properties of choice probabilities. A representative model of this approach was developed by Luce (1959). Based on an assumption referred to as the *"Choice Axiom"*, Luce developed the Multinomial Logit (MNL) model presented below. An alternative approach, called the *random utility approach*, was formulated by Manski (1977). According to this approach, the individual is assumed to select the alternative with the highest utility, but this utility is treated as a random variable. Under certain assumptions about the stochastic distribution of utilities, this approach also leads to the MNL model. The derivation of the MNL model and the equivalence of the two approaches are outlined in Appendix A.

The salient feature of the MNL model is that choice probabilities depend on the choice-set, or domain. Expanding the choice-set to include additional alternatives decreases the probabilities of all alternatives in the original set, providing positive probabilities to the additional alternatives.

In a population that consists of individuals with diverse preferences, all of them facing the same choice-set, this feature of the MNL model creates a natural tension: a larger choice-set which 'spreads' the choice probabilities over more alternatives, exacerbates the mistakes that some individuals make compared to perfect choice of their most preferred alternative. On the other hand, the larger choice-set may enable others to choose their more valued alternatives not included in the original set. This tradeoff suggests that there exists an optimum choice-set which maximizes social welfare, taking into account the distribution of preferences in the population. Specifically, the optimum choice-set is the subset of all alternatives which maximizes a utilitarian social welfare function whose elements are individual expected utilities. The objective of this paper is to explore the factors that affect the selection of the optimum choice-set.

A particularly attractive feature of the MNL model is that a certain parameter determines the 'spread' of choice probabilities among alternatives in the choice-set: high values of this parameter imply that individuals' choice probabilities are more concentrated around their highly valuable alternatives, choosing the best alternative with certainty in the limit. Small values of this parameter imply that individuals make close to purely random choice. This parameter can therefore be viewed as representing the precision of choice. We refer to it as the *'degree of rationality'* (with *'perfect rationality'* in the limiting case). We focus on the effect that the level of rationality has on the optimum choice-set. In two limiting cases, the optimum sets are the following (Section 2 below):

(a) At high degrees of rationality, all alternatives are included in the optimum choice-set;

(b) At low degrees of rationality, the optimum choice-set includes only one alternative, totally eliminating individual choice.

For intermediate degrees of rationality, the dependence of the optimum choice-set on the precision of choice is complex. We present numerical calculations that demonstrate that, for example, the standard single-peakednees assumption about individual preferences is insufficient to determine that the optimum choice-sets are nested and expand as the degree of rationality rises.

A more flexible policy than optimizing the number of alternatives included in the choice-set is to assume that the government can directly affect individuals' choice probabilities. We model this by assuming that the government can attach weights to the choice probabilities of different alternatives. In this context, the previous analysis of the optimum choice-set can be viewed as a special case when zero weights are attached to the choice probabilities of certain alternatives.

These weights reflect various means that governments have to affect individual choices. The importance, for example, of *'framing effects'* is well documented, particularly in the design of default options. See Johnson *et-al* (1993) on this effect in car insurance, Choi *et-al* (2003) and Caroll *et-al* (2009) on the dramatic effects of the design of opting-out and opting-in options on participation in 401(K) pension plans and on other insurance decisions, or Johnson and Goldstein (2003) on this effect in organ donations.

The determination of optimum probability weights is described in Section 4, followed by a binary example which highlights the dependence of these weights on individuals' degree of rationality (Section 5).

In Section 6 we apply the model to the early retirement eligibility constraint (the age at which individuals can start receiving retirement benefits) common to public social security systems (and private pension funds). Presumably, this constraint (which is age 62 in the US) strikes a balance between those who would sensibly like to retire earlier and those who would mistakenly retire too early, regretting this decision later in life. We analyze the critical degree of rationality at which it becomes socially desirable to impose this constraint in terms of the 'type one' and 'type two' errors that individuals make. It is shown that when more than 10-15 percent of the population choose mistakenly, then it is optimal to impose this constraint. This result, though, is highly sensitive to the level of the *'replacement rate'*, that is, the ratio of consumption in retirement to consumption during work.

Section 7 discusses briefly possible modifications required when the model is generalized to include heterogeneity in the degree of rationality. The basic message of this paper, though, is preserved: under bounded rationality, certain limits on individual choice are socially desirable.

There is a vast popular literature on "too much choice". Of particular interest are the books by Schwartz (2001) and the recent one by Iyengar (2010). Both books provide broad insights from psychology and sociology that, on the one hand, choice is a 'human condition' and, on the other hand, why consumers would be better-off with fewer options.

2 A Multinomial Logit Choice Model

We first specify the basic MNL model. This model will be amended later when we discuss policies that directly affect choice probabilities.

Consider a population consisting of heterogeneous individuals, each characterized by a parameter θ ("individual θ"). This parameter represents personal characteristics, such as health, longevity or attitudes towards work, which are regarded as private information. Individuals choose one among a finite number, n, of alternatives, numbered $i = 1, 2, .., n$. They attach a non-negative utility, $u_i(\theta)$, to each alternative. Choice is probabilistic. The MNL model specifies the probability that individual θ chooses alternative i, denoted $p_i(\theta, q)$, as

$$p_i(\theta, q) = \frac{e^{qu_i(\theta)}}{\sum_{j=1}^{n} e^{qu_j(\theta)}} \qquad i = 1, 2, .., n \tag{1}$$

where q is a positive parameter. Clearly, $0 \leq p_i(\theta, q) \leq 1$ and, since one and only one alternative is chosen, $\sum_{i=1}^{n} p_i(\theta, q) = 1$.

Note that the scale parameter q is assumed to be independent of θ. An interpretation of q as a parameter of the underlying random distribution from which (1) is derived (*the Gumbel distribution*), is presented in Appendix A. The role of q can best be understood by focusing initially on two limiting cases:

Case 1

$$p_i(\theta, 0) = \lim_{q \longrightarrow 0} p_i(\theta, q) = \frac{1}{n} \qquad i = 1, 2, .., n. \tag{2}$$

All alternatives are equally likely to be chosen;

Case 2

$$p_i(\theta, \infty) = \lim_{q \longrightarrow \infty} p_i(\theta, q) = \begin{cases} 1 & \text{if } u_i(\theta) > \max\limits_{\text{all } j \neq i} u_j(\theta) \\ 0 & \text{if } u_i(\theta) < \max\limits_{\text{all } j \neq i} u_j(\theta) \end{cases} \qquad i = 1, 2, .., n. \tag{3}$$

In the event of ties among the utilities of some alternatives, $u_i(\theta) = \max_{\text{all } j \neq i} u_j(\theta)$, the limit in case 2 is $\frac{1}{n^*}$, where $n^*(< n)$ is the number of alternatives for which $u_i(\theta) = \max_{\text{all } j \neq i} u_j(\theta)$, and is zero for the remaining $n - n^*$ alternatives. For simplicity we shall henceforth disregard ties.

It is seen that $q = \infty$ implies that individuals choose the best alternative, while $q = 0$ implies pure random choice independent of preferences. More generally, we shall show below that a higher q raises the probabilities of more valued alternatives and decreases the probabilities of less valued alternatives. Thus, q can be viewed as representing the *precision of choice*. At times we shall refer to q as the *degree of rationality* (with $q = \infty$ called *'perfect rationality'*).

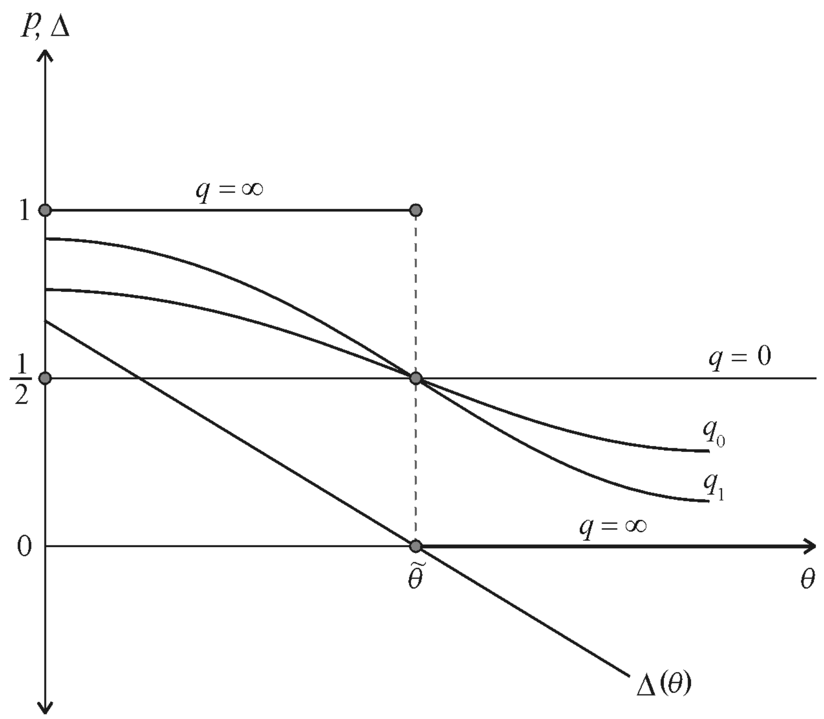

Figure 1 ($q_1 > q_0$)

Figure 1 portrays the MNL model for a binary example, $i = 1, 2$, with $\theta \geq 0$. Suppose that $\Delta(\theta) = u_1(\theta) - u_2(\theta) \gtreqless 0$ as $\theta \lesseqgtr \tilde{\theta}$ for some $\tilde{\theta} > 0$. Under perfect rationality, all individuals with $0 \leq \theta \leq \tilde{\theta}$ choose alternative 1, and those with $\theta > \tilde{\theta}$ choose alternative 2. Probabilistic choice leads to 'type one' and 'type two' errors:

some who would choose alternative 1 under perfect rationality actually choose alternative 2, and some who would choose alternative 2 under perfect rationality choose alternative 1.

One aspect of the MNL model which received much attention is the *Independence from Irrelevant Alternatives* (IIA): the ratio of any two probabilities i and j is unaffected by the utilities of other alternatives. From (1),

$$\frac{p_i(\theta)}{p_j(\theta)} = e^{q(u_i(\theta)-u_j(\theta))} \tag{4}$$

This property has important ramifications, yielding counterintuitive outcomes in some cases[1]. For policy purposes, however, as shown below, the IIA is an advantage for manipulations of individual choice probabilities aimed at raising social welfare.

3 Social Welfare and Individuals' Optimum Choice Set

Individuals' welfare is represented by expected utility, $V(\theta, q)$,

$$V(\theta, q) = \sum_{i=1}^{n} p_i(\theta, q) u_i(\theta) \tag{5}$$

A change in the parameter q affects V through the probabilities $p_i(\theta, q)$:

$$\frac{\partial p_i(\theta, q)}{\partial q} = p_i(\theta, q)(u_i(\theta) - V(\theta, q)) \quad i = 1, 2, .., n \tag{6}$$

[1] For example, the "Red-Bus/Blue-Bus Paradox" pointed-out by Debreu (1960). In the MNL model, equal probabilities ($\frac{1}{2}$) of going by car or by bus become, in a three-way choice between car, red bus and blue bus, equal probabilities of $\frac{1}{3}$ for each alternative. To avoid such paradoxes choice hierarchies have been offered: the IIA is applied first to choice between groups of 'similar' alternatives and then within these groups. However, finding an independent criterion for grouping alternatives may be problematic. A more general model ('simple scalability') proposed by Tversky (1972) avoids some of these questionable results, but has other problems (see Ben-Akiva and Lerman (1985)).

Since V is a weighted average of utilities, (5), a higher q is seen to raise the probabilities of alternatives whose utility is higher than average utility, V, and to decrease the probabilities of alternatives whose utility is lower than V.

Consequently, a higher q raises expected utility (omitting notation of the functional elements):

$$\frac{\partial V}{\partial q} = \sum_{i=1}^{n} \frac{\partial p_i}{\partial q} u_i$$

by (6)

$$= \sum_{i=1}^{n} p_i (u_i - V) u_i =$$

$$= \sum_{i=1}^{n} p_i (u_i - V)^2 > 0, \tag{7}$$

assuming that not all alternatives have the same utility.

Using (2) and (3), calculate the level of V in the limiting cases:

$$V(\theta, 0) = \lim_{q \to 0} V(\theta, q) = \frac{1}{n} \sum_{i=1}^{n} u_i \tag{8}$$

and

$$V(\theta, \infty) = \lim_{q \to \infty} V(\theta, q) = u_i(\theta) \tag{9}$$

where $u_i(\theta) > \max_{j \neq i} u_j(\theta)$.

When choice is purely random, expected utility is the (arithmetic) average of utilities. Under perfect rationality, individuals choose the alternative with the highest utility. Denote this maximum by $\bar{V}, \bar{V} = V(\theta, \infty)$.

Let social welfare, W, be utilitarian:

$$W(q) = \int_{\underline{\theta}}^{\bar{\theta}} V(\theta, q) dF(\theta) \tag{10}$$

where $F(\theta)$ is the distribution function of θ in the population. It is assumed that $F(\theta)$ is defined ever a finite, non-empty, interval, $(\underline{\theta}, \bar{\theta})$. For simplicity, we assume that the density $f(\theta) > 0$ exists for all $\theta \epsilon (\underline{\theta}, \bar{\theta})$. Since $V(\theta, q)$ strictly increases with q, so does $W(q)$, $\dfrac{dW(q)}{\partial q} > 0$.

(a) Optimum Choice-Sets for High and Low Degrees of Rationality

Individual utilities are private information, but the government knows the distribution of utilities in the population. Based on this distribution, the government is assumed to determine the set of alternatives from which individuals make choices, called the *'choice-set'*. In this context, optimum policy is taken to be the choice-set which maximizes social welfare.

Trivially, any alternative which has a lower utility than some other alternative for all θ's can be excluded from consideration since any choice-set which includes this alternative is inferior to some choice-set which excludes it. Consequently, we shall assume that no alternative is dominated by other alternatives for all θ's. This is formalized in the following assumption. For each alternative, define the set of θ's for which it is ranked highest by some individuals:

$$\Omega_i = \left\{ \theta \mid u_i(\theta) > \max_{j \neq i} u_j(\theta) \right\} \quad i = 1, 2, .., n \tag{11}$$

Assumption 1 *All* Ω_i, $i = 1, 2, .., n$ *are non-empty.*

Optimum policy for the two limiting cases discussed above is straightforward. Maximum social welfare, $\bar{W}$, is attained when all individuals choose their most preferred alternative. This can be attained when the choice-set includes *all* alternatives,

$$\bar{W} = \lim_{q \to \infty} W(q) = \int_{\underline{\theta}}^{\bar{\theta}} \bar{V}(\theta) dF(\theta) \tag{12}$$

Denote social welfare when all individuals choose alternative i, by W^i:

$$W^i = \int_{\underline{\theta}}^{\bar{\theta}} u_i(\theta) dF(\theta) \tag{13}$$

From (7),

$$W(0) = \lim_{q \to 0} W(q) = \frac{1}{n} \sum_{i=1}^{n} W^i \tag{14}$$

When all alternatives are included in the choice-set and individuals choose randomly, independent of preferences, social welfare is the arithmetic average of the W^i's.

These observations point-out the optimum choice-sets in the two limiting cases and, by continuity, for cases in their neighborhood:

Proposition 1 *For large and for small q's individuals' optimum choice-sets are as follows:*

(a) *When q is large, the choice-set includes all alternatives;*

(b) *When q is small, the choice-set is a singleton, that is, it contains one alternative, say alternative m, where $W^m > \max\limits_{j \neq m} W^j$.*

Part (a) follows directly from (9) and Assumption 1. Part (b) follows directly from (14).

Viewing parts (a) and (b) of Proposition 1, continuity of W in q implies that there exists a q_0, $q_0 > 0$, such that for all $q < q_0$, the optimum choice-set includes a single alternative, while for $q \geq q_0$ the choice-set includes two or more alternatives. Intuitively, it is expected that the optimum choice-set includes an increasing number of alternatives as q rises above q_0. We shall now study this process in some detail.

(b) Optimum Choice-Sets in Intermediate Cases

We now wish to address the general question of the dependence of the optimum choice-set on the level of q. Let S be a subset of $\{1, 2, 3, .., n\}$. The choice probabilities for $i \in S$,, $p_i^S(\theta, q)$, are equal to $p_i^S(\theta, q) = \frac{e^{qu_i(\theta)}}{\sum\limits_{i \epsilon S}^{n} e^{qu_i(\theta)}}$. By definition, the probabilities of alternatives not in S are zero $\left(\sum\limits_{i \epsilon S} p_i^S(\theta, q) = 1\right)$. With choice-set S, expected utility, $V^S(\theta, q)$, is equal to $V^S(\theta, q) = \sum\limits_{i \epsilon S} p_i^S(\theta, q) u_i(\theta)$, and social welfare, $W^S(q)$, is equal to $W^S(q) = \int\limits_{\underline{\theta}}^{\bar{\theta}} V^S(\theta, q) dF(\theta)$.

Definition 1 *The optimum choice-set for a given q, $S^*(q)$, is*

$$S^*(q) = \left\{ S \;\middle|\; \underset{S \,\leq\, \{1,2,..,n\}}{Max} W^S(q) \right\} \tag{15}$$

It was shown that $S^*(q)$ consists of a single alternative for small q's, while $S^*(q)$ consists of *all* alternatives for large q's. Characterization of $S^*(q)$ for intermediate values is complex. This can best be demonstrated by a numerical example.

Let there be three alternatives, each identified by a number x_i, $i = 1, 2, 3$, and three individuals θ_j, $j = 1, 2, 3$. Individual θ_j's utility of alternative i, u_i, is $u_i(\theta_j) = -(x_i - \theta_j)^2$. This formulation resembles Hotelling's (1929) model: individuals and stores (firms) are located on a line. Due to transportation costs, utility decreases with the distance of individual j's location, θ_j, from store x_i. In the calculations below:

θ_1	θ_2	θ_3
.5	1	2

x_1	x_2	x_3
1	.5	1.6

By construction, preferences are (Figure 2) *single-peaked* (e.g. Mass-Colell *et-al* (1995)).

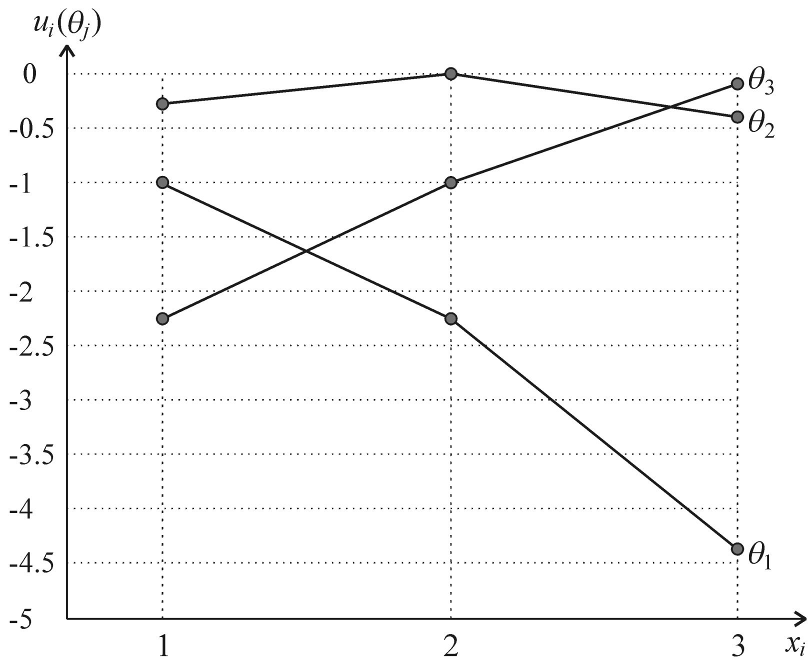

Figure 2

Possible choice-sets are $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{2,3\}$, $\{1,3\}$, and $\{1,2,3\}$. The corresponding social welfare levels are denoted W^1, W^2, W^3, $W^{1,2}$, $W^{2,3}$, $W^{1,3}$, and $W^{1,2,3}$, respectively. In Figure 3, each of these is plotted against different levels of q. For each q, the optimum choice-set is the outer envelope of these curves. The figure demonstrates Proposition 1: at low q's, the optimum set has a single alternative (W^1) and at high q's the optimum set includes all alternatives. Of particular interest is the fact that the optimum choice-sets are not nested as q rises. The choice-set $\{1,2\}$ is optimum for certain values of q between 2 and 3 while the set $\{2,3\}$ is optimum for still higher levels of q.

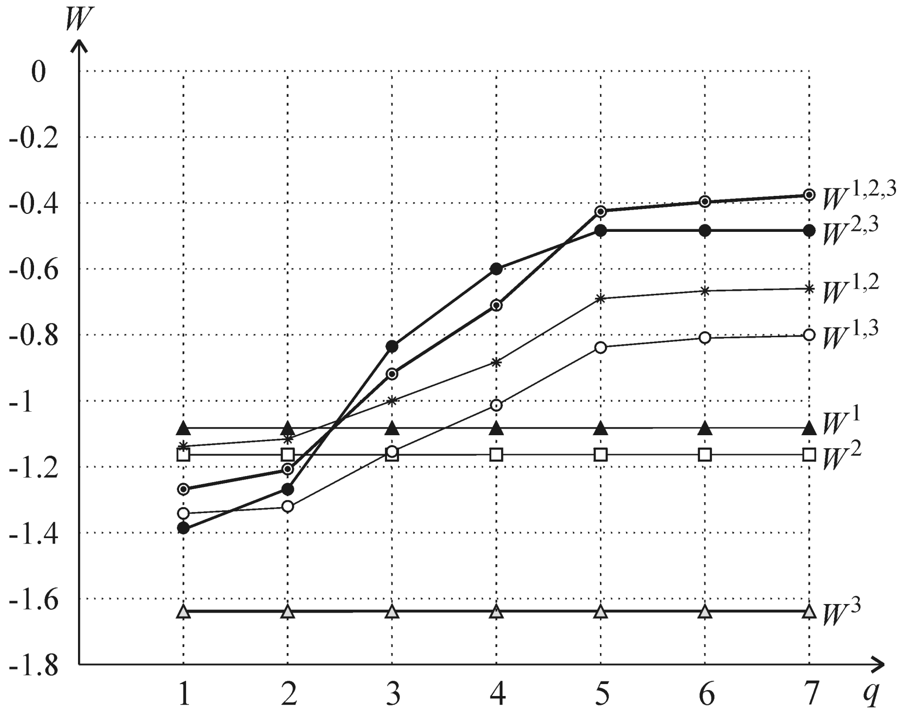

Figure 3

The above example also demonstrates that a subset of a certain set which has higher social welfare than the subset at some q may become socially superior at higher q's.

To understand the causes of such a switch, consider a subset $\widetilde{S}$ of S, $\widetilde{S} \subseteq S$. Using the above definitions,

$$
\begin{aligned}
W^S(q) - W^{\widetilde{S}}(q) &= \int_{\underline{\theta}}^{\bar{\theta}} (V^S(\theta,q) - V^{\widetilde{S}}(\theta,q))dF(\theta) = \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \sum_{i \in S-\widetilde{S}} p_i^S(\theta,q)(u_i(\theta) - V^{\widetilde{S}}(\theta,q))dF(\theta) = \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \left(\sum_{i \in S-\widetilde{S}} p_i^S \right) \left(V^{S-\widetilde{S}}(\theta,q) - V^{\widetilde{S}}(\theta,q) \right) dF(\theta) \qquad (16)
\end{aligned}
$$

where $S - \widetilde{S}$ is the set which contains all alternatives in S but not in $\widetilde{S}$, and $V^{S-\widetilde{S}} = \sum_{i \in S-\widetilde{S}} \left(\frac{p_i^S}{\sum_{i \in S-\widetilde{S}} p_i^S} \right) u_i$ is expected utility over the set of alternatives in $S-\widetilde{S}$. The integrand in (16) is the probability of choosing an alternative in $S-\widetilde{S}$ times net expected utility of the inclusion of $S - \widetilde{S}$ in the choice-set. The negative term $V^{\widetilde{S}}$ is the lower expected utility obtained from alternatives in $\widetilde{S}$, because the inclusion of the alternatives in $S - \widetilde{S}$ reduces the probability of choosing alternatives in $\widetilde{S}$. Suppose that $W^S - W^{\widetilde{S}} > 0$ for some q. Higher q's raise both $V^{S-\widetilde{S}}$ and $V^{\widetilde{S}}$ for all θ, reflecting the higher weight given to the most preferred alternative in each set. The probability of choosing any alternative in $S - \widetilde{S}$ increases for the highest utility in $S - \widetilde{S}$ and decreases all other probabilities. This lends higher weight to the increase in $V^{S-\widetilde{S}}$. Still $V^{\widetilde{S}}$ may rise more than $V^{S-\widetilde{S}}$ and even reverse the sign of $W^S - S^{\widetilde{S}}$. In the limit, though, the integrand in (16) is equal, for each θ, to the utility of the alternative in $S - \widetilde{S}$ with the highest utility, and this is clearly positive, consistent with Proposition 1.

4 Policy Affecting Choice Probabilities

Since individual preferences are assumed to be private information, policy can only distinguish between alternatives. In the previous sections it was assumed that the government determines the optimum choice-set available to individuals. This approach can be generalized by assuming that the government can directly affect individual choice probabilities. The previous discussion amounts to assignment of zero probability to certain alternatives, tantamount to exclusion from the choice-set.

The government has various ways to influence choice probabilities. The well documented tendency to choose default alternatives (e.g. Johnson *et-al* (1993) or Caroll *et-al* (2009)) is one example. Many studies show that "framing" issues (such as "opting-out" and "opting-in" design) affect individuals' choices (e.g. Choi *et-al* (2003)). These studies demonstrate that control over the method of choice enables the designer, whether the government or private firms, to affect choice probabilities. Governments may also use fiscal instruments to shift individuals' choice. As an example of the latter, consider the imposition of a tax/subsidy, t_i, on alternative i. The policy $\underline{t} = (t_1, t_2, .., t_n)$ affects the choice probabilities, (1), which are now rewritten

$$p_i(\theta, q, \underline{g}) = \frac{e^{q(u_i - t_i)}}{\sum\limits_{j=1}^{n} e^{q(u_j - t_j)}} = \frac{e^{qu_i} g_i}{\sum\limits_{j=1}^{n} e^{qu_j} g_j} \tag{17}$$

where $g_i = e^{-qt_i}$, $g_i \geq 0$, $i = 1, 2, .., n$ and $\underline{g} = (g_1, g_2, .., g_n)$. The vector $\underline{g}$ is the weights given to the choice probabilities. The government's objective is to choose the vector $\underline{g}$ that maximizes social welfare[2]. Of particular interest are cases when the optimum weight is zero, that is, when an alternative is excluded.

Note that $p_i(\theta, q, \underline{g})$ is homogeneous of degree zero in $\underline{g}$. One can therefore normalize $\sum\limits_{i=1}^{n} g_i = 1$.

[2] Another method to influence probabilities is the multi-stage choice process mentioned above. Such process can be shown to change in a predictable way the imputed choice probability of each alternative. To demonstrate this, consider three alternatives with utilities u_i, $i = 1, 2, 3, ...$ In a one-stage choice, the probability of choosing alternative i is $p_i = e^{qu_i} / \sum\limits_{j=1}^{3} e^{qu_j}$. Suppose that the choice is first between alternatives 1 and a 'package' consisting of alternatives 2 and 3. The expected utility of the 'package', denoted $\hat{u}_{23}$, is $\hat{u}_{23} = \hat{p}_2 u_2 + (1 - \hat{p}_2) u_3$ where $\hat{p}_2 = e^{qu_2} / (e^{qu_2} + e^{qu_3})$. The probability of choosing alternative 1 in a two-round choice is $\tilde{p}_1 = e^{qu_2} / (e^{qu_2} + e^{q\hat{u}_{2,3}})$. It is easy to show that $\tilde{p}_1 \leq p_1$, with strict inequality when there are no ties in the u_i's. Thus, the probabilities that each of the 'package' alternatives is chosen are raised in the two-stage process. This can be generalized to any finite numbers of alternatives in a multi-stage process.

The limiting cases (2) and (3) now become (as before, disregarding ties):

$$p_i(\theta, 0, \underline{g}) = \lim_{q \to 0} p_i(\theta, q, \underline{g}) = g_i \text{ and}$$

$$p_i(\theta, \infty, \underline{g}) = \lim_{q \to \infty} p_i(\theta, q, \underline{g}) = \begin{cases} g_i & \text{if } u_i(\theta) > \max\limits_{j \neq i} u_j(\theta) \\ 0 & \text{if } u_i(\theta) < \max\limits_{j \neq i} u_j(\theta) \end{cases} \tag{18}$$

As stated in Proposition 1, the optimum policy in the neighborhood of the limiting cases is clear. For large q's, that is, close to perfect rationality, all alternatives are included in the optimum choice-set, which means that $g_i > 0$, $i = 1, 2, .., n$ (when $q = \infty$, the specific $\underline{g}$ chosen is irrelevant). At low q's, close to pure random choice, the choice-set has only one alternative, say m, with $g_m = 1$ and $g_i = 0$ for all $i \neq m$. We want to explore the optimum weights, that is, the weights that maximize social welfare at some $q(\geq q_0)$. Denote these weights by $\underline{g}^*(q)$.

The F.O.C. are

$$\frac{\partial W(q, \underline{g}^*)}{\partial g_i} = \frac{1}{g_i^*} \int_{\underline{\theta}}^{\bar{\theta}} p_i(\theta, q, \underline{g}^*)(u_i(\theta) - V(\theta, q, \underline{g}^*) dF(\theta) \leq 0, \quad i = 1, 2, .., n \tag{19}$$

where $\frac{1}{g_i^*} p_i(\theta, q, \underline{g}^*) = \frac{e^{q u_i(\theta)}}{\sum\limits_{j=1}^{n} e^{q u_j(\theta)} g_j^*}$. Equation (22) holds with equality when $0 < g_i^* < 1.$,

The net benefit for individual θ from marginally increasing the weight of alternative i is the expected utility gain $\frac{1}{g^*} p_i u_i(\theta)$ due to the increase in the probability of this alternative minus the decrease in expected utility due to the decrease in the probabilities of all other alternatives.

We relegate the second-order conditions to Appendix B, stating there the conditions for uniqueness of $\underline{g}^*(q)$.

The determination of $\underline{g}^*(q)$ and the range of q's for which choice is eliminated, is brought-out clearly in the following binary example.

5 A Binary Example

Let $i = 1, 2,\ \ p = \dfrac{e^{qu_1} g}{e^{qu_1} g + e^{qu_2}(1-g)}$ is the probability of choosing alternative 1 and g, $0 \le g \le 1$, the weight given to this alternative. Using (19), the F.O.C. condition for the optimum g, g^*, is

$$\frac{\partial W(q, g^*)}{\partial g} = \int_{\underline{\theta}}^{\tilde{\theta}} \frac{e^{q\Delta(\theta)} \Delta(\theta)}{(e^{q\Delta(\theta)} g^* + 1 - g^*)^2} dF(\theta) \le 0 \tag{20}$$

with equality when $0 < g^* < 1$, where $\Delta(\theta) = u_1(\theta) - u_2(\theta)$. A necessary condition for an interior solution, g^*, is that Δ changes sign at least once over $(\underline{\theta}, \bar{\theta})$, that is, following Assumption 1, each of the two alternatives is ranked first by some individuals.

Let's assume that

$$\Delta(\theta) \gtreqless 0 \ \text{ as } \ \theta \lesseqgtr \tilde{\theta} \tag{21}$$

for some $\tilde{\theta}$, $\underline{\theta} < \tilde{\theta} < \bar{\theta}$. Then, for any $0 < g < 1$, $p(\theta, \infty, g) = 1$ for $\underline{\theta} < \theta < \tilde{\theta}$ and $p(\theta, \infty, g) = 0$ for $\tilde{\theta} < \theta < \bar{\theta}$. Maximum social welfare, $\bar{W}$, is

$$\bar{W} = W(\infty, g) = \int_{\underline{\theta}}^{\tilde{\theta}} u_1(\theta) dF(\theta) + \int_{\tilde{\theta}}^{\bar{\theta}} u_2(\theta) dF(\theta) \tag{22}$$

For the other limiting case,

$$W(0, g) = gW^1 + (1-g)W^2. \tag{23}$$

For concreteness, assume that $W^1 > W^2$, that is, when the choice-set has only one alternative, alternative 1 yields higher social welfare. Hence, by (19), when $q = 0$, the optimum policy is $g^* = 1$. In Figure 4, $W(0, g)$ is a linear function of g increasing from W^2 at $g = 0$ to W^1 at $g = 1$.

An increase in q raises W for all $0 < g < 1$. At $g = 1$, the slope $\dfrac{\partial W}{\partial g}$ decreases as q rises ($\dfrac{\partial}{\partial q}(\dfrac{\partial W(q,1)}{\partial g} = -\int_{\underline{\theta}}^{\bar{\theta}} e^{-q\Delta} \Delta^2 dF(\theta) < 0$) while at $g = 0$ this slope increases

$(\frac{\partial}{\partial q}(\frac{\partial W(q,0)}{\partial g} = \int_{\underline{\theta}}^{\bar{\theta}} e^{q\Delta}\Delta^2 dF(\theta) > 0)$. At any $q > q_0$, $\frac{\partial W(q,1)}{\partial g} < 0$, implying that there is a unique interior solution to (20), $0 < g^*(q) < 1$[3].

The direction of the change in $g^*(q)$ due to a small increase in q is indeterminate. It depends on the level of g^*: for high values of g^* an increase in q decreases g^*, and the opposite holds for small g^*'s.

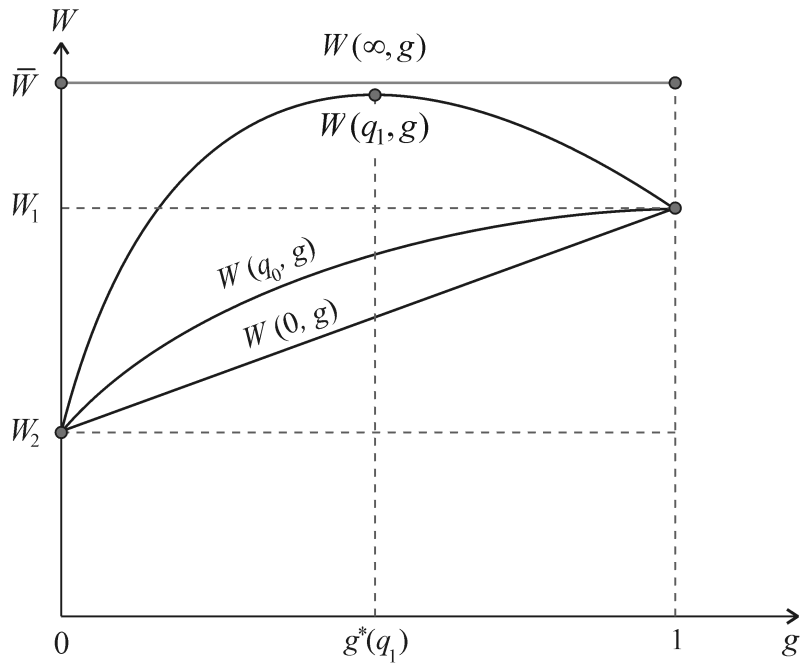

Figure 4 $(q_1 > q_0)$

We can use the binary example to further understand what factors affect q_0. Taking a linear approximation for $e^{-q\Delta}$, solve $\frac{\partial W(q_0,1)}{\partial g} = 0$ for q_0:

$$q_0 = \frac{W^1 - W^2}{\sigma_1^2 + \sigma_2^2 - Cov_{12}} \tag{24}$$

where $\sigma_i^2 = \int_{\underline{\theta}}^{\tilde{\theta}} (u_i(\theta) - W^i)^2 dF(\theta)$, $i = 1, 2$, and $Cov_{12} = \int_{\underline{\theta}}^{\tilde{\theta}} (u_1(\theta) - W^1)(u_2(\theta) - W^2) dF(\theta)$ are the variance of u_i, $i = 1, 2$, and the covariance of u_1 and u_2, respec-

[3] $\frac{\partial^2 W}{\partial g^2} = -2\int_{\underline{\theta}}^{\bar{\theta}} \frac{e^{q\Delta}\Delta(e^{q\Delta} - 1)}{((e^{q\Delta} - 1)g + 1)^3} dF(\theta) < 0$

tively. The higher is the mean preference for alternative 1 over 2, the higher is the threshold for introducing alternative 2. Conversely, the larger is the diversity of preferences, represented by the variances and (negatively) by the covariance, the lower is this threshold.

6 Application to Early Eligibility for Retirement Benefits

All public social security systems (and private pensions) have an early eligibility age at which a person can start receiving a pension. This age differs widely across countries. In the US it is age 62 and full benefits were reached at 65, moving gradually to 67. In the UK, early eligibility and full benefits are both at age 65. Imposing a constraint of earliest age for claiming benefits hurts workers who would 'sensibly' stop working before this age due to health and other personal circumstances. On the other hand, it prevents people from retiring too early due to inadequate savings or shortsightedness. The early eligibility age is supposed to strike a balance between these considerations. To obtain some idea about the magnitude of the mistakes which justify such a constraint, we shall apply a binary MNL model to choice between work and retirement.

Let $u(c_a)-\theta$ be workers utility where c_a is their consumption and θ is disutility from work. Let $v(c_b)$ be the utility of non-workers, where c_b is their consumption (pension benefits). Individuals differ in their labor disutility, θ, whose distribution in the population is $F(\theta)$. Take the range of θ to be $(0, \bar{\theta})$.

Under perfect rationality, individuals work or retire as $u(c_a)-\theta \gtreqless v(c_b)$. Define $\hat{\theta}$,

$$\hat{\theta} = \max(u(c_a) - v(c_b), 0) \tag{25}$$

Individuals with $\theta < \hat{\theta}$ work and those with $\theta > \hat{\theta}$ do not work (retire). Assume that $0 < \hat{\theta} < \bar{\theta}$, so that in the First-Best some work and some do not work.

With bounded rationality, the probability that a θ- individual works is given by

$$p(\theta, q) = \frac{e^{q(u(c_a)-\theta)}}{e^{q(u(c_a)-\theta)} + e^{qv(c_b)}} = \frac{e^{q(\hat{\theta}-\theta)}}{e^{q(\hat{\theta}-\theta)} + 1} \tag{26}$$

Denote social welfare when everyone works by $W_a = u(c_a) - E(\theta)$, where $E(\theta) = \int_0^{\bar{\theta}} \theta dF(\theta)$ is the expectation of labor disutility. Social welfare when nobody works is $v(c_b)$. We assume that everybody working is socially preferred to nobody working: $W_a > v(c_b)$. The relevant comparison is therefore between social welfare with a retirement choice, $W(q)$, and without the retirement option, W_a:

$$W(q) = \int_0^{\bar{\theta}} (p(\theta, q)(u(c_a) - \theta) + (1 - p(\theta, q))v(c_b))dF(\theta) \tag{27}$$

$$W^a = \int_0^{\bar{\theta}} (u(c_a) - \theta)dF(\theta) \tag{28}$$

Hence,

$$W(q) - W^a = -\int_0^{\bar{\theta}} (1 - p(\theta, q))(\hat{\theta} - \theta)dF(\theta) = -\int_0^{\bar{\theta}} \left(\frac{\hat{\theta} - \theta}{e^{q(\hat{\theta}-\theta)} + 1} \right) dF(\theta) \tag{29}$$

Since $W^a = \int_0^{\bar{\theta}} (\hat{\theta} - \theta)dF(\theta) + v(c_b)$ and, by assumption, $W^a - v(c_b) > 0$, it is seen that $W(0) - W^a < 0$. Retirement should not be an option when individuals decide whether to work or retire purely randomly.

As (29) strictly increases with q, there exists a $q_0 > 0$, defined by

$$\int_0^{\bar{\theta}} \frac{\hat{\theta} - \theta}{e^{q_0(\hat{\theta}-\theta)} + 1} dF(\theta) = 0 \tag{30}$$

such that for all $q > q_0$, having a retirement option is desirable.

We present in Table 1 calculations for the case $u(c) = v(c) = \ln c$ and $F(\theta)$ a uniform distribution over $(0, \frac{1}{3})$. Since $\hat{\theta} = \dfrac{u(c_a)}{v(c_b)} = \ln\left(\dfrac{c_a}{c_b}\right)$, we chose values for

$\frac{c_a}{c_b}$, the ratio of pre-retirement to post-retirement consumption (the inverse of the *'replacement rate'*), in the commonly observed range: 1.2, 1.25, and 1.3.

For each of these values we calculated $W^a - v(c_b) = \hat{\theta} - E(\theta) = \hat{\theta} - .056$. All these values are positive, as assumed.

Table 1

$\frac{c_a}{c_b}$	$\hat{\theta}$	q_0	*percent working*	E_1	E_2
1.2	.18	3.29	54	.24	.21
1.25	.22	13.10	67	.15	.12
1.3	.26	26.53	79	.09	.06

In Table 1, for each value of $\frac{c_a}{c_b}$, we calculate $\hat{\theta}\left(= \ln \frac{c_a}{c_b}\right)$, the percent of the population working in the First-Best ($= 3\hat{\theta}$) and q_0, the solution to (34).

Most insightful are the 'type one' and 'type two' errors at q_0, the level of q at which choice is introduced. That is, the percent of those who work in the First-Best but choose to retire under bounded rationality, and the percent of those who are non-workers in the First-Best but choose to work under bounded rationality. These are E_1 and E_2 in the last two columns of Table 1, defined as:

$$E_1(q_0) = \int_0^{\hat{\theta}} (1 - p(\theta, q_0))dF(\theta); \quad \text{and} \quad E_2(q_0) = \int_{\hat{\theta}}^{\bar{\theta}} p(\theta, q_0)dF(\theta) \qquad (31)$$

The size of these errors is seen to decrease significantly as $\frac{c_a}{c_b}$ increases. This is not surprising. A rise in this ratio raises the preference for work, decreasing the value of the non-work option.

It would be interesting to have some data about individuals' retrospective view about the extent that they misjudged their optimum retirement age, depending on their actual age of retirement. Of particular relevance are those who retire close to the early eligibility constraint.

7 Varying Degrees of Rationality Among Individuals

It has been assumed throughout that individuals have a common degree of rationality, q. Relaxing this assumption requires modification of certain conclusions. Suppose that individuals are identified by two parameters, θ and q. These parameters are assumed to be jointly distributed in the population. When the support of the (marginal) distribution of q is a narrow interval then the results in Proposition 1 are still applicable. Specifically, with small q's, the optimum choice-set is a singleton, and with large q's, all alternatives are contained in the optimum choice-set. However, when the support of the distribution of q is wide, that is, individuals have widely varying degrees of rationality, then some questions explored earlier have to be rephrased and conclusions modified.

Consider, for example, the following question: is there a fraction of individuals with high levels of q that warrants the inclusion of all alternatives in the optimum choice-set?

A binary choice model will demonstrate that the answer to this question is negative. Let choice be between two alternatives, 1 and 2. There are two types of individuals, each identified by the pair (θ_i, q_i), $i = 1, 2$. Let the fraction of type 1 individuals be f, $0 < f < 1$. Denote by $u_j^i = u_j(\theta_i)$, $i, j = 1, 2$, and $p_i(\theta_i, q_i)$ is the probability of type i individuals choosing alternative 1. Expected utilities, V^i, are $V^i = p^i u_1^i + (1 - p^i) u_2^i$, $i = 1, 2$, and social welfare, W, is $W(q_1, q_2) = V^1 f + V^2 (1 - f)$. When only alternative i is in the choice-set, social welfare is W^i, $W^i = u_i^1 f + u_i^2 (1 - f)$.

To have a meaningful problem, assume that the two types have opposite preferences: $\Delta^1 > 0$ and $\Delta^2 < 0$, where $\Delta^i = u_1^i - u_2^i$, $i = 1, 2$.

If $W^1 > W^2$, then alternative 1 is included in the choice-set for any (q_1, q_2). Starting with W^1, consider whether the inclusion of alternative 2 is desirable:

$$W(q_1, q_2) - W^1 = -\left(\frac{\Delta^1 f}{e^{q_1 \Delta^1} + 1} + \frac{\Delta^2 (1 - f)}{e^{q_2 \Delta^2} + 1} \right) \qquad (32)$$

The interpretation of (32) is straightforward. The first term is negative, being equal to the loss due to the fraction of type 1 individuals choosing alternative 2. The second term is positive and equal to the added utility of the fraction of type 2 individuals who choose their preferred alternative 2.

By assumption $W(0,0) - W^1 < 0$. To see the effect of large differences in the q's, take $q_2 = 0$. By (35),

$$W(q_1, 0) - W^1 = -\left(\frac{\Delta^1 f}{e^{q_1 \Delta^1} + 1} + \frac{\Delta^2 (1-f)}{2} \right) \tag{33}$$

It follows from (32) that $W(\infty, 0) - W^1 = \lim_{q_1 \to \infty} W(q_1, 0) - W^1 = -\frac{\Delta^2(1-f)}{2} > 0$. For large q_1, the choice-set includes both alternatives. Alternatively, let $q_1 = 0$. Then

$$W(0, q_2) - W^1 = -\left(\frac{\Delta^1 f}{2} + \frac{\Delta^2 (1-f)}{e^{q_2 \Delta^2} + 1} \right) \tag{34}$$

Now, $W(0, \infty) - W^1 = \lim_{q_2 \to \infty} W(0, q_2) - W^1 = -\left(\frac{\Delta^1 f}{2} + \Delta^2 (1-f) \right) \gtreqless 0$.

Even when all type 2 individuals choose perfectly, alternative 2 may not be included in the optimum choice-set.

The difference between the limiting cases (33) and (34) is that in the former case, the welfare loss due to inclusion of alternative 2 becomes negligible relative to the welfare gain. In the latter case, the loss is finite and so is the highest gain (when all type 2 individuals choose alternative 2).

As this example demonstrates, when a fraction of the population has a high degree of rationality this does not suffice for inclusion of all alternatives in the optimum choice-set. In this sense, the desirability of imposing limits on choice remains the defining feature of our model.

Appendix A

We describe two alternative approaches to derive the MNL model.

Constant Utility Approach

The decision maker chooses probabilistically, with utilities as parameters. This approach makes specific assumptions about the structure of the probabilities. In the text, $p_i^S(\theta, q)$ denotes the probability that the θ-individual chooses alternative i from the choice-set S, where $q > 0$, a constant, plays a central role in the analysis. Here we omit the notation for θ and q. Focusing on the choice-set, we write $p_i(S)$ to be the probability of choosing alternative i from the set S, $i \in S$. Let $\tilde{S}$ be a subset of S, $\tilde{S} \subseteq S$. The probability of choosing subset $\tilde{S}$ from S, $\tilde{S} \subseteq S$, $p_{\tilde{S}}(S)$, is equal to $p_{\tilde{S}}(S) = \sum\limits_{i \in S} p_i(S)$. The *Choice-Axiom* formulated by Luce (1959) makes the following assumption about the probabilities:

A set of choice probabilities defined for all subsets of a finite set S, satisfy the choice axiom provided that for all i, $\tilde{S}$ and S, such that $i \in \tilde{S} \subseteq S$,

$$p_i(S) = p_i(\tilde{S}) p_{\tilde{S}}(S) \tag{A.1}$$

In words, the choice probability of i from S is the product of the choice probability of i from a subset $\tilde{S}$ that contains i and the probability that the choice lies in $\tilde{S}$. This is assumed to hold for any subset which contains i. (A.1) implies the property of *independence of irrelevant alternatives* (IIA):

$$\frac{p_i(\tilde{S})}{p_j(\tilde{S})} = \frac{p_i(S)}{p_j(S)}, \quad i \neq j, \quad i, j \quad \in \tilde{S} \subseteq S \tag{A.2}$$

Luce argues that (A.2) can be viewed as a probabilistic version of the property of transitivity. However, the *"Red bus - Blue bus Paradox"* discussed above (f.n.1) demonstrates that the choice-axiom is less appropriate when alternatives are similar and may, in turn, induce manipulations of the choice-set.

Luce (1959) proves that if the choice-axiom holds and a positive utility measure, U_i, is proportional to the probability of i, then these probabilities can be written

$$p_i(S) = \frac{U_i}{\sum_{j \in S} U_i} \tag{A.3}$$

The utilities U_i are unique up to multiplication by a positive constant (see the discussion in Ben-Akiva and Lerman (1985)). Writing $u_i = \ln U_i$, (A.3) is seen to be the MNL model

$$p_i(S) = \frac{e^{u_i}}{\sum_{j \in S} e^{u_j}} \tag{A.4}$$

Random Utility Approach

Manski (1977) made the argument that while individuals always choose the alternative with the highest utility, these utilities are not known to the analyst with certainty and should therefore be treated as random variables. For a description of the possible sources of randomness, such as unobserved attributes (θ in our notation), see Manski (1977) and Ben-Akiva and Lerman (1985).

The probability that alternative i will be chosen is equal to the probability that its utility, U_i, is greater or equal than the utilities of all other alternatives in S:

$$p_i(S) = pr\{U_i \geq U_j, \quad \text{all} \quad j \in S\}, \quad i = 1, 2, .., n \tag{A.5}$$

(disregarding ties). Specific assumptions on the joint distribution of the random utilities $\{U_i, \ i \in S\}$ are required in order to solve (A.5).

Manski assumes that utilities have a deterministic ("systematic") component, which can in principle be estimated, denoted V_i, and a pure disturbance term, denoted ε_i:

$$U_i = V_i + \varepsilon_i, \quad i = 1, 2, .., n \tag{A.6}$$

Hence,

$$p_i(S) = pr\{V_i + \varepsilon_i \geq \max_{\substack{j \in S \\ j \neq i}} (V_j + \varepsilon_j)\} \tag{A.7}$$

In order to solve (A.7), assumptions are made on the joint distribution of $\{\varepsilon_1, \varepsilon_2, .., \varepsilon_n\}$. Generally, this is quite complex. However, when all the disturbances are $i.i.d$s and follow a *Gumbel distribution* with a common scale parameter q, $q > 0$, then (A.7) is equivalent to the MNL model (see Ben-Akiva and Lerman (1985)):

$$p_i(S) = \frac{e^{qV_i}}{\sum\limits_{j \in S} e^{qV_j}} \tag{A.8}$$

The *Gumbel distribution*, $F(\varepsilon)$, is

$$F(\varepsilon) = \exp[-e^{-q\varepsilon}] \tag{A.9}$$

This distribution has the properties that any linear transformation of the ε's is also Gumbel distributed, the difference between two Gumbel distributed variables, $\varepsilon_1 - \varepsilon_2$, is also Gumbel distributed and, most important, when ε_i are all i.i.d. Gumbel distributed, then the $\max(\varepsilon_1, \varepsilon_2, .., \varepsilon_n)$ is also Gumbel distributed. These properties enable the derivation of (A.8). For the original derivation see Domencich and McFadden (1975). Ben-Akiva and Lerman (1985) discuss the importance of the i.i.d. assumption and homoscedasticity.

Appendix B

Optimum Probability Weights

The F.O.C., (19) in the text, are

$$\frac{\partial W(q,\underline{g}^*)}{\partial g_i} = \frac{1}{g_i^*}\int_{\underline{\theta}}^{\bar{\theta}} p_i(\theta,q,\underline{g}^*)\Delta_i(\theta,q,\underline{g}^*)\ dF(\theta) \leq 0, \quad i=1,2,..,n \tag{B.1}$$

where $\Delta_i(\theta,q) = u_i(\theta) - V(\theta,q,\underline{g}^*)$, $i = 1,2,..,n$. Equation (B.1) with equality when $0 < g_i^* < 1$.

Using the definition of $p_i(\theta,q,\underline{g})$, (20), after some manipulations,

$$\frac{\partial^2 W(q,\underline{g}^*)}{\partial g_j\ \partial g_i} = -\frac{1}{g_i^*\ g_j^*}\int_{\underline{\theta}}^{\bar{\theta}} p_i(\theta,q,\underline{g}^*)p_j(\theta,q,\underline{g}^*)(\Delta_i(\theta,q,\underline{g}^*)+$$
$$+\Delta_j(\theta,q,\underline{g}^*))dF(\theta) \quad i,j=1,2,..,n \tag{B.2}$$

Sufficient second-order conditions are that the matrix $\left[\frac{\partial^2 W}{\partial g_i\ \partial g_j}\right]$ is negative semi-definite (because of the constant $\Sigma g_i = 1$).

Diagonal terms are equal to

$$\frac{\partial^2 W(q,\underline{g}^*)}{\partial g_i^2} = -\frac{2}{g_i^{*2}}\int_{\underline{\theta}}^{\bar{\theta}} p_i^2(\theta,q,\underline{g}^*)\Delta_i(\theta,q,\underline{g}^*)\ dF(\theta) \tag{B.3}$$

Assume that $p_i\Delta_i$ is monotone in θ, changing sign at some $\tilde{\theta}_i$, $\underline{\theta} < \tilde{\theta}_i < \bar{\theta}$. Suppose that $p_i\Delta_i$ decreases in θ (same argument applies in the opposite case). Since $p_i > 0$ also decreases in θ,

$$p_i^2(\theta,q,\underline{g}^*)\Delta_i(\theta,q,\underline{g}^*) > p_i(\tilde{\theta},q,\underline{g}^*)p_i(\theta,q,\underline{g}^*)\Delta_i(\theta,q,\underline{g}^*) \tag{B.4}$$

for all $\theta \in (\underline{\theta},\bar{\theta})$. Integrating, using (B.1) (for an interior $\underline{g}^*$):

$$\int_{\underline{\theta}}^{\bar{\theta}} p_i^2(\theta, q, \underline{g}^*)\Delta_i(\theta, q, \underline{g}^*)dF(\theta) > p_i(\tilde{\theta}, q, \underline{g}^*)\int_{\underline{\theta}}^{\bar{\theta}} p_i(\theta, q, \underline{g}^*)\Delta_i(\theta, q, \underline{g}^*)dF(\theta) = 0 \tag{B.5}$$

Hence, by (B.3), $\dfrac{\partial^2 W}{\partial g_i^2} < 0, i = 1, 2, .., n$.

A sufficient condition for an off-diagonal element to be positive is that $p_i(\theta, q, \underline{g}^*)$ and $p_j(\theta, q, \underline{g}^*)$ change with θ in opposite directions for all $i \neq j$. Thus, when $p_i\Delta_i$ decreases in θ and $p_j\Delta_j$ increases in θ, both changing sign once over $(\underline{\theta}, \bar{\theta})$ at $\tilde{\theta}_i$ and $\tilde{\theta}_j$ respectively, then (omitting the notation of the elements in the functions), at an interior $\underline{g}^*$, rewrite (B.2)

$$\begin{aligned} \frac{\partial^2 W}{\partial g_i\, \partial g_j} &= -\frac{1}{g_i^*\ g_j^*}\int_{\underline{\theta}}^{\bar{\theta}} p_i p_j(\Delta_i + \Delta_j)dF(\theta) = \\ &= -\frac{1}{g_i^*\ g_j^*}\left[\int_{\underline{\theta}}^{\bar{\theta}} p_j(p_i\Delta_i)dF(\theta) + \int_{\underline{\theta}}^{\bar{\theta}} p_i(p_j\Delta_j)dF(\theta)\right] \end{aligned} \tag{B.6}$$

Each term in the square brackets satisfies $p_j(p_i\Delta_i) < p_j(\tilde{\theta}_i, q, \underline{g}^*) \cdot p_i\Delta_i$, for all θ. Integrating, using (B.1)

$$\int_{\underline{\theta}}^{\bar{\theta}} p_j(p_i\Delta_i)dF(\theta) < p_j(\tilde{\theta}_i, q, \underline{g}^*)\int_{\underline{\theta}}^{\bar{\theta}} p_i\Delta_i\ dF(\theta) = 0 \tag{B.7}$$

The same calculation shows that the second term in the square brackets is negative. It now follows from (B.6) that $\dfrac{\partial^2 W}{\partial g_i\, \partial g_j} > 0$.

Of course, for more than two alternatives, the above argument cannot apply to *all* pairs i, j, $i \neq j$. We have to assume directly that all off-diagonal elements in the matrix $\left[\dfrac{\partial^2 W}{\partial g_i\, \partial g_j}\right]$ are positive. This is well-known to imply that this matrix is negative semi-definite.

References

[1] Ben-Akiva, M. and S.Lerman (1985), *Discrete Choice Analysis* (MIT).

[2] Caroll, G., J.Choi, D.Laibson, B.Madrian and A.Metrick (2009), "Optimal Defaults and Active Decisions", *Quarterly Journal of Economics*, 121, 505-540.

[3] Choi, J., D.Laibson and B.Madrian (2003) "Optimal Defaults", *American Economic Review, Papers and Proceedings*, 93, 180-185.

[4] Debreu, G. (1960), "Review of D.Luce Individual Choice Behavior. A Theoretical Analysis", *American Economic Review*, 50, 186-188.

[5] Domencich, T. and D.McFadden (1975) *Urban Travel Demand: A Behavioral Analysis* (North Holland).

[6] Hotelling, H. (1929), "Stability in Competition", *The Economic Journal*, 39, 41-57.

[7] Iyengar, S. (2010), *The Art of Choosing* (Twelve).

[8] Johnson, E. and D.Goldstein (2003), "Do Defaults Save Lives?", *Science*, 1338-1339.

[9] Johnson, E. J.Hershey, J.Mezaros and H.Kunrenther (1993), "Framing, Probability Distortions and Insurance Decisions", *Journal of Risk and Uncertainty*, 7, 35-51.

[10] Luce, D. (1959), *Individual Choice Behavior: A Theoretical Analysis* (Wiley).

[11] Luce, D. and P.Suppes (1965) "Preference, Utility and Subjective Probability" in *Handbook of Mathematical Psychology*, vol.3, R.Luce, R.Bush and E.Galanter, eds. (Wiley).

[12] Manski, C. (1977), "The Structure of Random Utility Models", *Theory and Decision*, 8, 229-254.

[13] Mass-Colell, A., M.Whinston and J.Green (1995), *Microeconomic Theory* (Oxford).

[14] Schwartz, B. (2004), *The Paradox of Choice: Why More is Less* (Harper).