Why do I keep going in circles?
On the impossibility of constructing rational preferences

Brendan Markey-Towler[*†]

3rd March 2014

Abstract

In standard models of rational choice it is typically taken for granted that preferences are given and defined over the alternatives alone, and the possibility of making a rational choice is simply a matter of assumption. In this paper I generalise this aspect of the economic model so that preferences over alternatives are constructed from given preferences defined over various characteristics of the alternatives under consideration. I characterise the decision problem before investigating what conditions a procedure for aggregating preferences over attributes into preferences over alternatives must satisfy in order for the latter to be rational. I then consider what the implications of these conditions for the procedural rationality of the aggregation process.

Keywords: Decision making, multiple attributes, aggregation, possibility, rationality
JEL classification: C00, D01, D11, D21

1 Choice with multiple attribute alternatives

The possibility of rational choice lies at the very heart of modern economic theories of behaviour, the existence of a well behaved preference relation guiding this choice providing the foundation for nearly every standard model of economic behaviour. Indeed, even the increasingly prominent "behavioural" economics research program which attempts to incorporate insights from psychology into economics takes rational choice to be the fundamental principle guiding behaviour (see Rabin, 2013a,b). Typically, we simply assume that such

[*]University of Queensland, School of Economics, contact email: brendan.markeytowler@uqconnect.edu.au
[†]The author thanks Simon Grant for extensive and invaluable discussions, as well as Jacqueline Robinson for comments on early drafts of the present work

choice is possible by constructing "given" binary preferences between alternatives under consideration for any individual which conform to the axioms of completeness and transitivity or specifying a utility function representing them (see Mas-Collel et al., 1995, Ch.2). The preference between the alternatives is taken to be given, primitive to the model, a fact.

However, in constructing such binary preferences between alternatives, we are making a significant implicit assumption. We are assuming consumers judge the *gestalt* of the alternatives they must choose from, considering the alternatives as a whole rather than judging them by a comparison of their characteristics. This is a significant assumption to make because in an objective sense (i.e. external to the individual), there is nothing other to distinguish the alternatives from one another than their name, no characteristics which they have which could explain why one is different to another other than the individual's preference. A more subtle interpretation of the assumption implicit in taking a preference between alternatives as given and primitive is that we are assuming that consumers can always aggregate the comparison of the attributes of the alternatives into a well-behaved rational preference relation defined over alternatives. But even this somewhat semantic interpretation is unjustified unless we can demonstrate that consumers can always meaningfully aggregate considerations of multiple attributes into preferences over alternatives, and that therefore we can safely ignore considerations of the attributes characterising the alternatives.

The assumptions implicit in taking preferences between alternatives to be primitive were abandoned briefly during the 1960s by economists developing the now somewhat neglected "New Consumer Theory". The foundations of this theory were laid out in ground-breaking work by Lancaster (1966) and Ironmonger (1972), who realised that what really distinguishes one alternative from another is the differentiation of their characteristics and the preferences over these characteristics. Alternatives are judged on the basis of the attributes which characterise them, and so the preferences for alternatives will be determined by the preferences for their basket of attributes. The new consumer theorists incorporated this idea by defining utility functions over "characteristics" spaces and thus over the alternatives themselves via a "technology" representing the characterisation of alternatives by points in that space.

The great beauty of the "New Consumer Theory" was that it provided a means by which the new goods and services generated by a constantly evolving economy could be incorporated into the preference structure of the individual in a fairly systematic manner. If a standard approach is taken to preference then any theory of demand in an evolutionary economy must make a bald assumption about the integration of new alternatives into the preferences of individuals, plain and simple. For instance, in the theory of "growing awareness" of new elements of state space, von Neumann-Morgenstern utility (which represent rational preferences) can be extended to the new state space domain under fairly innocuous

assumptions (Karni and Viero, 2013), but there is no theory concerning the value the extended utility function assigns to those new elements. In the New Consumer Theory, the only assumption required for a new alternative to be incorporated into the preference structure is that it be characterised by attributes. This gives it a location in characteristic space which is then transformed into a location in the preference structure by the pre-existing procedure for transforming preferences over attributes into preferences over alternatives. Hence the theory surrounding the idea that preferences over alternatives arises from preferences over their attributes is of vital importance to any theory of demand in an evolutionary economics. It gives us a theoretical means of determining consumer demand for novel products other than plain assumption.

However, in defining utility functions over "characteristics" spaces, and keeping the theoretical discourse in the Euclidean space realm of utility functions which impose rational preference relations by the transitivity of the real numbers Lancaster (1966), Ironmonger (1972) and (famously) Stigler and Becker (1977) imposed possibility of constructing rational preferences represented by a utility function over alternatives from the utilities of attributes without investigating whether this was justifiable on a preference-axiomatic level. Nonetheless, the new consumer theorists were correct in realising that if it is indeed the differentiation of attributes which distinguish alternatives, then it must be a comparison of these attributes which at least in part constitutes comparison of the alternatives themselves. Hence, preferences over alternatives must be constructed from preferences over attributes.

If we are to accept that preferences over alternatives are not in themselves primitive but arise from more deeply primitive given preferences defined over the *attributes* which characterise those alternatives, we cannot *prima facie* say that it is always possible to construct a rational preference relation over the alternatives themselves. This problem was examined in an interesting paper, Kapeller et al. (2012), in which the authors gave a simple, plausible example where the aggregation of otherwise rational preferences over attributes of alternatives aggregates to irrational preferences over the alternatives themselves. Suppose we had three alternatives and three characteristics as represented in Table 1, where attributes can take one of four values, in decreasing order of preference, $\{++, +, \circ, -\}$

	x_1	x_2	x_3
a_1	$\circ$	$++$	$+$
a_2	$\circ$	$-$	$+$
a_3	$+$	$\circ$	$-$

Table 1: A characterisation of alternatives $x \in X$ (adapted from Kapeller et al., 2012, p.43)

Now, if we give equal weight to each of the characteristics in the aggregation to preference

over alternatives, we can see that x_1 is preferred to x_2, and x_2 is preferred to x_3, but x_3 is preferred to x_1, so rational choice in the sense of choosing the most preferred outcome is impossible given the spread of attributes across the characteristics space and the aggregation scheme employed. Thus there are cases of choice in the presence of multiple attributes where rational choice is impossible.

While I have demonstrated elsewhere[1] that Kapeller et al. (2012) erroneously draw a mathematical equivalence between this problem and Arrow's impossibility theorem (by providing a counter-example where rational choice *is* possible), their example serves to show that the possibility of rational choice in the presence of multiple attribute alternatives is not a forgone conclusion and requires examination. The question I endeavour to begin to answer here is; under what conditions is rational choice in the presence of multiple attribute alternatives possible?

The concern with taking account of multiple attributes in the process of constructing preferences and making choices places the problem of choice at the nexus of "procedural rationality" and "substantive rationality". Simon (1976) distinguished "procedural" rationality from "substantive" rationality, which concerns itself whether individuals are making a choice consistent with a given objective (in economics, the maximisation of utility). Substantive rationality in the context of multiple attribute alternatives therefore concerns the possibility of constructing a rational preference relation between alternatives which can guide the choice amongst alternatives.

Procedural rationality by contrast concerns the process of thought in which an individual engages before coming ultimately to a decision, and asks whether this process is reasonable not only given the basic objectives of the individual, but also given the constraints imposed by cognitive ability. Of particular relevance to the procedural rationality then are the computational demands placed on the individual by the problem (see Simon (1976; 1978b; 1978a)). In the problem under consideration here, where the procedure of relevance is that by which preferences are constructed from preferences over attributes, procedural rationality concerns the complexity and intricacy of the procedure by which these preferences are arrived at. It should be stated from the outset however, that the notion of procedural rationality is not as formally clear cut as that of the (economic) notion of substantive rationality by virtue of the fact that it concerns human cognitive ability, so it is difficult to make definitive absolute statements of a "this is procedurally rational solution" nature. Hence the discussion of procedural rationality here will be conducted in informal relative terms.

So in making an effort to answer the question of when rational choices are possible in the context of multiple attribute alternatives, we are asking what conditions must be

[1] Markey-Towler (2014)

satisfied in order to find a *process* by which a substantively rational preference relation can be constructed. Obviously, the conditions required for the choice environment and the process itself will have implications for the procedural rationality of any solution to the choice problem. Perhaps unsurprisingly, I will argue that these conditions imply something of a trade-off between substantive and procedural rationality.

In the next section I formally define the choice problem in the context of multiple attribute alternatives on the level of preference relations. With this in hand, I can then begin to make some headway into the question of under what conditions rational choice is possible in such scenarios, providing two rather trivial theorems on conditions which guarantee a rational choice always exists, before providing some more informative and interesting necessary and sufficient conditions for the existence of rational choice. I conclude with a discussion of the meaning of these conditions, and how they relate to the possibility of meaningful substantively rational choice when alternatives are characterised by their attributes in a constantly evolving system such as an economy. It will be argued that constructing a substantively rational preference relation may require either the use of the theoretically problematic class of lexicographic procedures (which do not admit utility representation), or such involved procedures for aggregating attribute preferences that substantive rationality comes at the price of procedural irrationality.

2 The choice problem

In any choice problem the most basic of primitives is a set X of alternatives amongst which a choice must be made. I will assume that this set is finite for simplicity[2]. In economics it is standard to suppose that choice is then guided by a preference relation $\succeq$ which is defined over the set X, such that there exists a set $\succeq(X)$ consisting of the elements of X pre-ordered according to $\succeq$. I will assume throughout that $|X| \geq 3$ so that rationality is never an economically trivial and uninteresting matter of specifying a single binary preference. While it is perfectly possible that such preferences relate only to the *gestalt* of the alternative at hand, as I have suggested above, it is equally possible that each alternative has various attributes which serve to distinguish it in a more objective sense from the others and which factor into the determination of $\succeq$. If this is indeed the case then each alternative under consideration will be characterised by its attributes.

Generalising the concept of attribute spaces suggested by Lancaster (1966) and Iron-

[2]The alternatives set could be extended to incorporate infinitely many alternatives, though additional technical assumptions would be required (compactness) to guarantee the existence of a maximal element, which would complicate the exposition here for little gain in generality.

monger (1972), if alternatives x are characterised by their attributes then there exists an arbitrary characteristic space $A = \Pi_{i=1}^{|A|} A_i$ within which the attributes of x are defined. This space is taken to consist of the cross-product of multiple subspaces which divide A into "classes" represented by one subspace A_i indexed by $\{1, \ldots, |A|\}$, each of which we could say represents a particular "characteristic". Hence each alternative is distinguished from the others by its properties defined within each subspace. One could say $a_i \in A_i$ defines an alternative's attribute "for that particular characteristic".

If we were to go further and specify that A were a collection of metric subspaces, such that characteristic attributes were quantitative and contained within the set of real numbers $\mathbb{R}$, we could think of these attribute subspaces as *dimensions* of the characteristic space A, each dimension representing the quantities an alternative may have of a particular characteristic. However, for what follows I will leave A to be an arbitrary subspace[3].

Formally, we can now define an alternative to be characterised by the set of its attributes, $x = \{a_i\}_{i=1}^{|A|}$, organised into a vector $\{a_i\}_{i=1}^{|A|} \in \Pi_{i=1}^{|A|} A_i$ within the characteristic space. We can reverse this definition and define $a_i(x)$ to be the attribute of x within the characteristic subspace A_i. Implicit in this specification is the assumption, carried throughout what follows, that every alternative x is well defined in every characteristic subspace A_i by an attribute $a_i(x)$. This is largely a philosophical distinction that the lack of a particular characteristic in itself constitutes an attribute and hence a point within a characteristic subspace[4], and while it represents something of an antimony, it is mathematically important.

If $A = \Pi_{i=1}^{|A|} A_i$ is a space which characterises alternatives X it follows that an individual will have preferences$\succeq_i$ defined over each characteristic subspace A_i in much the same manner as over the alternatives space, so that we have a set of attribute preferences $\{\succeq_i\}_{i=1}^{|A|}$ defined over each individual attribute subspace $A_i \subset A$ which pre-order the characteristics space A such that there exists a set $\Pi_{i=1}^{|A|} \{\succeq_i (A_i)\}$ of pre-ordered attribute vectors. These preferences I take to be much the same as standard preferences and to have the same interpretation such that we can read $a_i(x) \succeq_i a_i(x')$ as "attribute a_i of x is at least as preferred as attribute a_i of x'". It is important here to note that an economically strong assumption has been made in specifying each of these preferences $\succeq_i$ over A_i alone, this assumption being that preferences over a particular characteristic subspace are defined independently of attributes in *other* subspaces. This means that if a particular alternative x performs particularly strongly on a particular attribute $a_i(x)$, the preferences $\succeq_{-i}$ over the rest of its attributes $\{a_{-i}(x)\}$ are not affected. I am of the belief that this assumption, while strong, is justifiable on the grounds

[3]We need no restrictions on finiteness here because we do not require a maximal element to exist in sets which merely serve to characterise points in another set

[4]Again, if we were to take $A_i \subset \mathbb{R}$, the implicit assumption I make here is that $0 \in A_i$ so that having zero "quantity" of a particular characteristic is the attribute 0

(other than simplicity) that we are not particularly interested in the attribute preferences *per se*, but rather the manner in which they contribute to preferences over alternatives. Moreover, the aggregation process can go some way toward mitigating the affect of this strong assumption by varying the importance of certain regions of the space A_i in the determination of $\succeq$, which is in keeping with the idea that preferences over attributes are defined for the attributes alone, and these preferences are used to construct preferences over alternatives.

Now, to obtain a well-defined solution to the problem of choice amongst alternatives in the presence of multiple attribute alternatives which accounts for those attributes, we require some means of aggregating preferences over each attribute subspace A_i into preferences over the alternatives space X. This is done by a *process,* described by a mapping $f : \Pi_{i=1}^{|A|} \{\succeq_i (A_i)\} \rightarrow \succeq (X)$, which takes the pre-ordered characteristic subspaces and transforms them into a pre-ordered space of alternatives characterised by their attributes within those spaces. This process involves giving a "weight" or "importance" or "priority" to the preferences over attributes in the characteristics subspaces $\succeq_i$ in the determination of preferences $\succeq$ over the alternatives which they characterise[5]. Thus f represents part of the process by which a choice will be arrived at, and in a sense can be thought to describe a thought process by which an individual constructs their preferences. This is a special type of the broad class of "procedures" for thinking about a particular problem of choice which Herbert Simon (1976; 1978b; 1978a) named as the subject matter of the study of "procedural rationality"[6].

An important and economically strong assumption will be made concerning the process f. I take f to be defined across each pre-ordered subspace $\succeq_i (A_i)$ alone, and invariant for all points within this subspace, so that each attribute space is given a certain importance in the determination of preferences. This imposes an ordinal interpretation on preferences and avoids thorny questions of cardinality of preferences within subspaces and across them, so that the individual is simply asking "do I prefer x or x' with respect to characteristic A_i?", "is characteristic A_i more important than A_j?" and so on and so forth rather than "*by how much* do I prefer this attribute to that attribute?" and "*how much* more important is this characteristic than that?"[7].

This is a controversial assumption, but one whose validity I cannot hope to convince anyone to accept, given the long-standing dispute between the Hicks-Samuelson, ordinal-

[5]I do not mean this to exclude the possibility any f which specifies a lexicographic construction of preference with respect to attributes.

[6]The study of procedural rationality is distinct from the study of "procedural utility", which defines hedonic utility functions over some quantitative representation of the process by which an outcome is arrived at Benz (2007), (Frey, 2008, pp.107-126).

[7]A possible avenue for future research would be to relax this assumption so that weights can vary across the possible attributes within the subspace A_i reflecting the idea that in certain regions of the attribute subspace A_i that particular subspace may be of less importance than in other regions.

ist interpretation of preference and the cardinalist interpretation (enjoying a resurgence in happiness economics). Suffice to say, I find it difficult to imagine a positive theory with an verifiable, universal, non-constructed[8] measure concept for comparing attributes within characteristic spaces and between them on a cardinal level, given that standard utility function representation of preferences is by definition constructed from revealed preference, and we do not observe directly (at present) the hypothesised "utility" generated by such choices. Indeed, the interpretation of a utility function as a cardinal representation of preference relies on the acceptance of a hotly contested 19th century psychological philosophy of Jeremy Bentham that all pleasures are ultimately reducible to a single hedonistic measure, no matter their source. In the multiple attribute setting, such an assumption would correspond to assuming a cardinal utility function defined not only over each individual attribute subspace, but also a cardinal utility function taking these utilities as arguments, f being contained within the supra-utility function such as in the "utility tree" manner of Strotz (1957), though Strotz' utility trees ultimately have roots not in attributes but in alternatives classed according to type.

Cardinal utilitarianism was challenged on philosophical grounds by Bentham's own student and founding political economist John Stuart Mill (1863) who argued that certain satisfactions are simply incomparable. The question remains far from settled in favour of a cardinalist interpretation of preference from a scientific standpoint as well. Indeed, the wealth of scientific evidence seems to point the other way, that many desires (and by extension the attributes from which they are derived) *are* non-comparable. In psychology, the theory of hierarchical needs (Maslow, 1943) presupposes that the nature of satisfaction depends on the need it is derived from, while any neural science textbook (such as Kandel et al. (2013)) will demonstrate that there are *many* pleasurable chemicals secreted within the brain, each pleasurable in their own way (dopamine, serotonin, adrenaline, testosterone being just a few). It seems far less strong to me, and to be on safer scientific grounds to presume only that preference is of a strictly "I prefer this to that" and "this characteristic is more important than that" nature rather than add an additional assumption and hypothesise the existence of some cardinal aggregation and hedonic scales.

Taken together, the objects $\{X \quad A \quad f\}$ make up the primitives of the choice problem, and more specifically specify the primitives of a procedure for aggregating preferences over the attributes of various alternatives under consideration into preferences over the alternatives themselves. Choice then consists of selecting an alternative x from X according to some

[8]Of course, all measures are constructed in a sense. I mean constructed in the sense that the object itself being measured is not directly observable and quantifiable in the physical world but is a hypothesised phenomenon of the mind

criterion. If this choice is to be "rational" in the substantive, economic sense (Simon, 1976), the theory states that the choice, x^*, will be the maximally preferred element in X, or $x^* = \max_{x \in X} \succeq (X)$. However, for this solution to be well-defined (as either a point-valued or set-valued solution) requires that maximal elements exist, and for this to be true of all sets within the power set 2^X we require the preference relation $\succeq$ be rational. If this were not true of all sets within 2^X then there would be a subset of alternatives $Y \subset X$ for which no $\succeq$ maximal elements exist due to intransitivity.

Following standard conventions in economics[9], we say that any preference relation $\succeq$ or $\succeq_i$ is rational if it is complete and transitive, that is, either $x \succeq x'$ or $x \preceq x'$ or both $\forall x, x' \in X$ and for any $x, x', x'' \in X$, $x \succeq x' \,\&\, x' \succeq x'' \implies x \succeq x''$ respectively. But given that $\succeq$ is emerging from a procedure f aggregating attribute preferences$\{\succeq_i\}_{i=1}^{|A|}$ it is also interesting whether the resultant preference relation is a "trivial" aggregation of these sub-preferences. By a "trivial" preference relation I mean one where a single characteristic subspace has *absolute* supremacy in the determination of the aggregated preference relation over the alternatives characterised by those attributes, so that preferences over alternatives are entirely determined by preferences over attributes in that characteristic subspace. Formally, a *non*-trivial preference relation can be defined thus:

Definition 1. A preference relation is *non-trivial* if and only if

$$\nexists i \in \{1, \ldots, |A|\} : x \succeq x' \iff a_i(x) \succeq_i a_i(x')$$

In a sense, a trivial preference relation is similar (but *not* equivalent as Kapeller et al. (2012) argue) to a dictatorial social choice rule in that one preference relation out of many determines the overall preference between alternatives. When a preference relation $\succeq$ is trivial in this manner, one characteristic subspace is dictatorial in the construction of preference between alternatives. But such preference relations $\succeq$ are not particularly interesting in the context of multiple attributes, as the full characteristic space is not taken into account in any meaningful way in the procedure f, given that the preference relation on alternatives is determined by only one of the characteristics of those alternatives. There is little difference in such cases between choice based on the *gestalt* of the alternative and choice based on consideration of attributes. Choice in the presence of then multiple attributes becomes a distinction without a difference. Hence we can define a class of preference relations of interest according to the following criterion:

Definition 2. A preference relation $\succeq$ constructed by f from preferences over attributes $\{\succeq_i\}_{i=1}^{|A|}$ is *non-trivially rational* if and only if

[9] See Mas-Collel et al. (1995, p.6) or Jehle and Reny (2011)

$$
\begin{array}{ll}
(1) & \forall x, x' \in X \text{ either } x \succeq x' \text{ or } x' \succeq x \text{ or both} \\
(2) & x \succeq x' \text{ and } x' \succeq x'' \text{ imply that } x \succeq x'' \\
(3) & \nexists i \in \{1, ..., |A|\} : x \succeq x' \iff a_i(x) \succeq_i a_i(x')
\end{array}
$$

The question of central importance then is: under what conditions will rational choice be impossible in any non-trivial sense due to the non-existence of a procedure f which can construct a rational preference relation over alternatives $\succeq$? Or, to put the point more bluntly, under what conditions can we find a procedure f such that substantively rational choice will be possible? What are the requirements upon the underlying preferences over attributes $\{\succeq_i\}_{i=1}^{|A|}$ and the procedure f by which these preferences are aggregated for the resulting preference relation defined over the alternatives to be non-trivially rational?

To make the problem more tractable again, I will assume for what follows that the pre-orderings $\{\succeq_i\}_{i=1}^{|A|}$ of the attribute spaces $\{A_i\}_{i=1}^{|A|}$ are rational, hence complete and transitive. This is not as strong an assumption as might appear at first glance and it gives us a good starting point for analysing under what conditions underlying attribute preferences aggregate to preferences over alternatives. If we were to have monotonic preferences over a quantitative attribute space A_i (not an entirely unrealistic assumption *ceteris paribus*) then the resulting preference relation will be transitive, and if this space were quantised (discrete) and had few points not much information would be required for the preferences to be complete. With this, we are now investigating under what conditions the rationality of underlying preferences will be maintained in the aggregation process.

3 Possibility theorems

For my purposes it will be useful to construct some additional concepts and their formal definitions before presenting theorems on the requirements on f and $\{\succeq_i\}_{i=1}^{|A|}$ for $\succeq$ to be non-trivially rational. Obviously if f consists of a weighting system for the dimensions on which the attributes characterising the alternatives and preferences over these are defined then the set of dimensions on which an alternative dominates another becomes critical. Hence we can define a set $S_x(x')$ consisting of a collection of the characteristic subspaces for which the attributes of alternative x weakly dominate those of x'.

Definition 3. The set of attribute dimensions for which x is at least as preferred as x' is given by $S_x(x') = \{A_i \subset A : a_i(x) \succeq_i a_i(x')\}$.

Note that with the traditional definition of $\succeq$ (see Mas-Collel et al., 1995, p.7), if $A_i \in S_x(x') \cap S_{x'}(x)$ then the alternatives are indifferent within that subspace and $A_i \in I_x(x') = \{A_i \subset A : a_i(x) \sim_i a_i(x')\} \subset S_x(x')$.

As per the choice problem outlined above this classification of A according to $\{\succeq_i\}_{i=1}^{|A|}$ must be aggregated by f into $\succeq$, and this requires that f give assign dominance of preferences over certain characteristic subspaces over others in the aggregation of attribute preferences to determine preference over the alternatives. Informally, a set of attribute subspaces can be said to dominate another when agreement of preferences between the attributes of two alternatives in the former determines the preference between the two alternatives. This dominance relation I will denote as $d.(f)$, so that we say $A'd.(f)A''$ reads as "$A' = \{A_i \in A'\} \subset A$ dominates $A'' = \{A_i \in A''\} \subset A$ in the aggregation by procedure f of preferences over attributes into preferences over alternatives". Formally dominance is defined as follows

Definition 4. For any two mutually exclusive collections of characteristic subspaces $A', A'' \subset A$, $A' \subset A$ dominates $A'' \subset A$ when agreement of preferences within the first set determines the preference between alternatives, that is, if $A' = S_x(x')$ and A'' then $x \succeq x'$ if and only if $A'd.(f)A''$.

Note that non-triviality can then be restated according to this definition, for if it is the case that no single preference relation over attributes has absolute supremacy in the determination of preferences over alternatives all the others it must be the case that no one attribute subspace dominates all the others in f, nor any combination of them. Given that f is in effect a weighting procedure, it follows that if one attribute subspace dominates all others in the aggregation it will dominate any single one of them also. Hence we can say that, $\succeq$ is non-trivial if and only if $\nexists A_i \subset A : A_i d.(f)\{A \setminus A_i\}$

We now have convenient notation which helps in establishing three preliminary, but fundamental, theorems providing conditions under which a rational preference over alternatives can be constructed from rational preferences over their attributes. The first two almost need no proof and simply confirm trivial conditions under which the preference over alternatives will be rational. However, given that the preference relation $\succeq$ is no longer given but constructed from given underlying preferences they will be stated for the sake of completeness. The first confirms that trivial preference relations will be rational, while the second confirms that if the spread of attributes is "trivial" in the sense that the attribute preferences between any two alternatives agree then the preferences over all alternatives will be rational.

Theorem 5. *If $\succeq$ is trivial such that $\exists i \in \{1 \quad |A|\} : x \succeq x' \iff a_i(x) \succeq_i a_i(x')$ (there is an absolutely dominant attribute subspace), then $\succeq$ will be rational.*

Proof. If $\exists A_i \subset A : A_i d.(f)\{A \setminus A_i\}$ then preferences over A_i absolutely dominate preferences over all other alternatives taken together, and hence $\succeq$ is completely determined by $\succeq_i$ and $\{a_i(x)\}_{x \in X}$, so since we have assumed that all alternatives are well defined in all

characteristic subspaces, for any two alternatives either $a_i(x) \succeq_i a_i(x')$ or $a_i(x') \succeq_i a_i(x)$ or both, and accordingly $x \succeq x'$ or $x' \succeq x$ or both. Similarly, suppose we have for any three alternatives $a_i(x) \succeq_i a_i(x')$ and $a_i(x') \succeq_i a_i(x'')$, so that by the definition of dominance, $x \succeq x'$ and $x' \succeq x''$. Since $\succeq_i$ is rational by assumption it follows that $a_i(x) \succeq a_i(x'')$ and thus by the triviality of $\succeq$ that $x \succeq x''$. □

This is a slightly more important result than its "triviality" would suggest. It demonstrates that as long as the preferences across attributes are rational, it is *always* possible to find a procedure with which a rational preference relation can be constructed when considering the attributes of alternatives.

Remark 6. Rational choice in the presence of multiple attribute alternatives is always possible if we do not exclude trivial procedures for determining preferences $\succeq$.

However, such a preference relation over alternatives is not interesting insofar as the guarantee of rationality comes at the cost of triviality in the process of considering the attributes of the alternatives. If this is indeed the case and we have a preference relation $\succeq$ that is trivial, then there is no meaningful consideration of multiple attributes, as in effect the alternative is equated with its attributes in the "dictatorial" characteristic subspace.

If we are to consider a wider range of aggregations f than trivial ones, then what conditions guarantee that the preference over alternatives will be rational? This is the content of the second "trivial" theorem concerning rational choice in the presence of multiple attribute alternatives.

Theorem 7. *If $\forall x, x' \in X$ all attribute preferences "agree" such that $a_i(x) \succeq_i a_i(x') \forall i \in \{1, \ldots, |A|\}$ or $a_i(x) \preceq_i a_i(x') \forall i \in \{1, \ldots, |A|\}$ or both, then $\succeq$ will be rational.*

Proof. Arbitrarily select any two alternatives $x, x' \in X$. Completeness of $\succeq$ follows immediately from the fact that for any two $x, x' \in X$ either $a_i(x) \succeq_i a_i(x') \forall i \in \{1, \ldots, |A|\}$ or $a_i(x) \preceq_i a_i(x') \forall i \in \{1, \ldots, |A|\}$ or both due to the agreement of attribute preferences, and so for *any* aggregation scheme f either $x \succeq x'$ or $x \preceq x'$ or both. Now without loss of generality assume that $a_i(x) \succeq_i a_i(x') \forall i \in \{1, \ldots, |A|\}$ so that $x \succeq x'$ for any aggregation scheme f. Since it was assumed above that $|X| \geq 3$ there must exist another $x'' \in X$. Let us suppose we have selected $x, x', x'' \in X$ so that for any aggregation scheme f we have $x' \succeq x''$, which, since attributes are all in agreement for any two alternatives across all characteristic subspaces, implies that $a_i(x') \succeq_i a_i(x'') \forall i \in \{1, \ldots, |A|\}$. It follows from the transitivity of $\succeq_i \forall i \in \{1, \ldots, |A|\}$ that $a_i(x) \succeq_i a_i(x'') \forall i \in \{1, \ldots, |A|\}$ and so $x \succeq x''$, and the preference relation $\succeq$ is transitive. Thus for any aggregation f we have a rational preference relation $\succeq$. □

Note that Theorem 7 implies that if the alternatives under consideration are characterised by attributes such that all the preferences between those alternative's attributes agree then we can guarantee that a rational preference relation over alternatives exists independently of the procedure f for aggregating attribute preferences. Indeed, we can construct a rational preference relation $\succeq$ with *any* aggregation scheme f provided that *all* characteristic subspaces agree on the preference between the attributes of two alternatives. This result suggests that the possibility of rational choice will depend in part on variables which lie beyond the individuals' control, the position of the various alternatives in the characteristic space A, and the degree of agreement of preferences between alternative attributes across the various subspaces of A.

While these theorems hint towards the requirements for rational choice they are still trivial in that extreme conditions on either the aggregation mapping f or the attributes $\left\{\{a_i(x)\}_{i=1}^{|A|}\right\}_{x\in X}$ are required to *guarantee* the existence of a rational preference relation over alternatives. They provide unsatisfactorily extreme sufficient conditions for the possibility of aggregating $\{\succeq_i\}_{i=1}^{|A|}$ into $\succeq$. What they serve to demonstrate is that the possibility of rational choice will depend on the choice of the particular procedure f for aggregating attribute preferences (under the individuals' control) as well as the positions of the various alternatives in the attribute space A (outside the individuals' control). It is to providing necessary and sufficient conditions on these objects for $\succeq$ to be rational that I now turn.

Theorem 8. *Define $A' = \{A_i \in A'\} \subset A$ to be a collection of subspaces of A. Suppose that for some arbitrary alternatives $x, x', x'' \in X$ we have a non-trivial aggregation mapping f such that $x \succeq x'$ and $x \succeq x''$. Then $x'' \succ x$ (a violation of transitivity) if and only if*

$$\exists A' \neq \emptyset : \begin{cases} A_i \notin S_x(x') \cap S_{x'}(x'') \;\forall\, A_i \in A' & (1) \\ \qquad \& \, A' d.\,(f)\, S_x(x'') & (2) \end{cases}$$

Intuitively, conditions (1) and (2) together prohibit the existence of a set of dominant characteristic subspaces which do not conform to a "linear" (in the non-functional sense) pattern in the spread of attributes through characteristic subspace.

Proof. Sufficiency: It follows from the transitivity of $\succeq_i$ that $a_i(x) \succeq a_i(x')$ and $a_i(x') \succeq a_i(x'')$ implies $a_i(x) \succeq_i a_i(x'')$. Then, by the definition of $S_x(x')$ and $S_{x'}(x'')$ as the sets of attribute subspaces in which one alternative dominates another, for any $A_{-i} \in S_x(x') \cap S_{x'}(x'')$ it is true that $A_{-i} \in S_x(x'')$, because

$$S_x(x') \cap S_{x'}(x'') = \{A_i : a_i(x) \succeq a_i(x')\} \cap \{A_i : a_i(x') \succeq a_i(x'')\}$$

$$\Longrightarrow S_x(x') \cap S_{x'}(x'') = \left\{ A_i : \begin{array}{c} a_i(x) \succeq a_i(x') \\ \& \, a_i(x') \succeq a_i(x'') \end{array} \right\}$$

and transitivity therefore implies

$$S_x(x') \cap S_{x'}(x'') = \{A_i : a_i(x) \succeq a_i(x'')\} = S_x(x'')$$

so it is the case that $A_i \notin S_x(x'')$. Now, since each preference relation for attribute subspaces is complete, and there is one defined for each subspace, we have that $S_x(x'') \cup S_{x''}(x) = A$. So because $A_i \notin S_x(x'')$, it follows not only that $A_i \in S_{x''}(x)$, but that $A_i \in S_{x''}(x) \setminus I_{x''}(x)$ and so $a_i(x'') \succ a_i(x)$. Hence if $\exists A' \neq \emptyset : A_i \notin S_x(x') \cap S_{x'}(x'') \; \forall A_i \in A'$, this implies that $\exists A' \neq \emptyset : A_i \in S_{x''}(x) \setminus I_{x''}(x) \; \forall A_i \in A'$, and then by the definition of the dominance relation $d.(f)$ and the content of the set A', $A'd.(f) S_x(x'') \Longrightarrow x'' \succ x$.

Necessity: Suppose that we have a non-trivial aggregation mapping f such that for some arbitrary alternatives $x \succeq x'$ and $x \succeq x''$ and $x'' \succ x$. Now, by the definition of the dominance relation $d.(f)$, and the definition of $S.(\cdot)$, $x'' \succ x$ only if $S_{x''}(x) \setminus I_{x''}(x) \, d.(f) \, S_x(x'')$. But by the argument employed in the proof of sufficiency above, $S_x(x'') = S_x(x') \cap S_{x'}(x'')$, so $x'' \succ x$ presupposes the existence of a non-empty collection of attribute subspaces $A' \subset A$ such that $A_i \notin S_x(x'') = S_x(x') \cap S_{x'}(x'') \; \forall A_i \in A'$ and $A'd.(f) S_x(x'')$. □

By demonstrating for any arbitrarily selected alternatives what is required of the aggregation mapping and the attributes of the alternatives under consideration to lead to intransitive preferences over alternatives, this Theorem provides necessary and sufficient conditions for rational choice to be possible in the multiple attribute alternative environment. Completeness is here a trivial aspect of the problem due to the completeness of the underlying preferences vis-a-vis attributes, which guarantees that a binary preference can always be constructed. If irrationality is to be present in the preference relation $\succeq$ then it must be due to a violation of transitivity, the conditions for which are outlined by Theorem 8. Hence we have an possibility result which can be summarised thus:

Corollary 9. *Non-trivial rational choice out of a set of three or more alternatives is possible if and only if conditions (1) and (2) of Theorem 8 do not hold for any three alternatives selected arbitrarily from X.*

Proof. Rational choice is defined as selecting the maximally preferred element, $x^* = \max_{x \in X} \succeq (X)$, but, given that X is countably finite, this problem is only well defined if the conditions of Theorem 8 fail to hold. □

4 Discussion

The conditions which are necessary and sufficient for substantively rational choice to be possible as I have written them may seem rather abstract and economically meaningless. However they merely say that for rational choice which considers in a non-trivial way the attributes of the alternatives, we need the spread of the attributes of alternatives in characteristic space to be somewhat "linear", so that preferences between attributes tend towards agreement between any three alternatives. On a more intuitive level, non-trivial rational choice requires that alternatives tend to perform either strong or weak across a sufficiently broad range of characteristics, rather than performing strongly in a certain class and weakly on others.

Non-trivial rational choice, where the consideration of attributes is meaningful, is therefore possible if and only if the process f by which they are considered gives sufficiently little weight to characteristic subspaces where alternatives which are dominated on other characteristic subspaces are performing strongly. And as more alternatives which are otherwise dominated become strong in a particular set of characteristic subspaces, this mapping must give increasingly little weight to those subspaces to maintain transitivity at the level of alternatives. Hence to preserve substantive rationality the procedure f must become increasingly "dictatorial" with respect to attribute subspaces which "go against the grain" of the general spread of attributes, such as it were.

It can be seen then that in the presence of multiple attribute alternatives the procedure which constructs the required preference for non-trivial rational choice can become quite intricate, as an schema f must be found which preserves completeness and transitivity over *all* alternatives. While by Theorem 5 it is *always* possible to construct a rational preference if we include the possibility of trivial "dictatorial" aggregation schemas independently, if a non-dictatorial aggregation is to be found then the possibility of rational choice depends in large part on the nature of the differentiation of the alternatives under consideration along the pre-ordered subsets of attributes. As I have shown above, this requires a significant degree of "transitivity" across many subsets of the attributes of the alternatives under choice, or potentially a highly selective and intricate procedure for aggregating these attributes which conforms to conditions (1) and (2) of Theorem 8 for *every* triad of alternatives. It seems reasonable to assert that as the procedure f becomes more and more intricate, the task of constructing preferences over alternatives from preferences over attributes becomes computationally more complex, and hence more cognitively taxing.

This is important in an economic setting, particularly when the problem of choice is that of the consumer, or buyer in general. In economies with profit-seeking producers, it is rational

for producers to differentiate their products in order to generate profits from market power. Profit-seeking is the driver of evolutionary dynamics in the economy, with differentiation of products (alternatives from which to choose) being the cause of variety generation (or origination) on which replication-selection dynamics (or self-organisational dynamics) can operate through consumer choice (Dopfer et al., 2004). Producers in such an economy *ceteris paribus* should seek to differentiate their products by augmenting their attributes in subspaces where competitors are weak rather than strong, "shifting the goalposts" in order to create market power rather than "playing catch-up", trying to make their products at least as good as their competitor's and decrease their competitor's market power with no real great gain to their own. It is generally more profitable to create a new, differentiated product than imitate an established one.

However, such tendencies on the part of producers would make the existence of transitive agreement of alternatives across a broad range of their attribute subspaces a rather remote possibility, and hence reduce the possibility of rational consumer choice without a potentially computationally demanding procedure for aggregating preferences over attributes into preferences over alternatives. If this is indeed the case, then substantive rationality comes at the price of sacrificing procedural rationality, for as Simon (1976; 1978b; 1978a) argues, procedural rationality places a premium on the computational simplicity of the process of thought which the cognitively constrained individual must engage in. So while substantively rational choice may be possible, the procedure which supports it may be procedurally irrational due to its placing a computational demand on the individual made impossible by the cognitive and time constraints of human beings.

Now, when faced with a highly differentiated set of alternatives amongst which to make a choice, and given the implication of Theorem 8 that procedurally irrational mappings f may be required to maintain substantive rationality, the results above suggest two possible avenues which will guarantee a substantively rational choice can be made. First, Theorem 5 gives rise to the option of selecting a trivial f which gives absolute priority to one characteristic for the determination of preferences over alternatives. Since the properties of the attribute preferences will follow to alternatives preferences, substantively rational choice will be possible, but as I have argued above (p.2), trivial preferences are uninteresting to a theory of multiple attribute alternatives because they do not consider multiple attributes.

The second avenue for guaranteeing substantively rational preferences is to apply a lexicographic procedure to attribute preferences. I have been careful thus far to state only that a *potentially* computationally demanding procedure f might be required to aggregate the attribute preferences of highly differentiated alternatives to guarantee substantive rationality. As a matter of fact, lexicographic preferences, it is well known, always generate a

rational preference relation and are a very simple class of procedures to implement once a lexicographic ordering has been arranged. However, lexicographic procedures for constructing preferences are extremely problematic for economic theory, because if preferences $\succeq$ are determined by a lexicographic procedure, it is well known that they cannot be represented by a utility function (see Mas-Collel et al. (1995, p.46) or Rubinstein (2006, p.15)). Hence the rules of differential calculus would no longer be available to represent choice elegantly as the solution of an optimisation problem.

Lexicographic preferences constitute an ordinal aggregation mapping f which can be specified as follows

$$f = \left\{ x \succeq x' \iff \begin{array}{ll} & a_1(x) \succeq_1 a_1(x') \\ \vee & a_1(x) \sim_1 a_1(x') \wedge a_2(x) \succeq_2 a_2(x') \\ \vdots & \vdots \\ \vee & \wedge_{i=1}^{|A|-1} \{a_i(x) \sim_i a_i(x')\} \wedge a_{|A|}(x) \succeq_{|A|} a_{|A|}(x') \end{array} \right\}$$

so clearly dominance of any one attribute subspace over another depends on its position in the index set, specifically, $A_i d.(f) A_j \iff i < j$, and given the lexicographic nature of the procedure, it is unnecessary to think of dominance in any other than a binary manner. Using this definition, we can recast the proof of the rationality of lexicographic preferences in terms of Theorem 8 but more importantly demonstrate their non-triviality:

Theorem 10. *Lexicographic preferences are non-trivially rational*

Proof. Non-triviality: Suppose that we have $k > 1 : a_k(x') \succ_k a_k(x)$ and that $a_i(x) \sim_i a_i(x') \; \forall i < k$. Now by the standard conceptualisation of $\succeq$, $a_i(x) \succeq_i a_i(x') \; \forall i < k$, but $x' \succ x$ by the definition of $\{\succeq(X)\} = f\left(\{\succeq_i (A_i)\}_{i=1}^{|A|}\right)$ so since k was arbitrarily chosen amongst the dimensions dominated by attribute subspace A_1 and $x' \not\sim x$, lexicographic preferences assign no absolute priority to a single attribute subspace.

Rationality: The proof of completeness is trivial, simply note that by the definition of f, as long as a pair of attributes exists within every attribute subspace (which is the case by assumption), then the preferences over the alternatives they characterise is well-defined. But for $\succeq$ to be rational also requires transitivity. Suppose that f is lexicographic. Then $A_i d.(f) A_j \iff i < j$. Now if $x \succeq x' \,\&\, x' \succeq x''$ then by definition of f it must be the case that $\exists A_i \in S_x(x') : A_i d.(f) S_{x'}(x)$ and $\exists A_j \in S_{x'}(x'') : A_j d.(f) S_{x''}(x')$. But since $x \succeq x'$, $i < j$ must be the case, so $k > i \forall A_k \in S_{x''}(x') \cup S_{x'}(x)$. Hence by the definition of f, $A_i d.(f) A_k \forall A_k \in S_{x''}(x') \cup S_{x'}(x)$. Now, if (1) of Theorem 8 holds, then any $A_n \in A'$ is not in $S_x(x') \cap S_{x'}(x'')$. But notice that $\neg\{S_x(x') \cap S_{x'}(x'')\} \subset S_{x'}(x) \cup S_{x''}(x')$, so any $A_n \in A'$ is also in $S_{x'}(x) \cup S_{x''}(x')$ and therefore dominated by both A_i and A_j. Hence under

lexicographic preferences, if (1) of Theorem 8 holds, then (2) does not, and preferences are rational. □

This is an interesting result because it implies that since lexicographic preferences are non-trivial, non-trivially rational preferences are always possible, and thus that rational choice is always possible in mathematically possible:

Remark 11. Non-trivial rational choice in the presence of multiple attribute alternatives is always possible if a lexicographic procedure f is adopted.

Given that lexicographic procedures are also computationally quite efficient, requiring the same complex considerations as would a "checklist", this could partly explain why they are widely observed empirically in the psychological literature on decision theory (Earl, 1990), as they provide a computationally efficient manner in which a substantively rational decision can be made.

Lexicographic preferences are a special class of procedurally rational heuristics which can be applied to the problem of choice. In general, these procedures are represented by simple mappings $g : A \to C(X)$ between the characteristic space A into a choice set $C(x)$, which need by no means be a singleton for any particular realisation $g(a \in A)$. The best known of these procedures is the simple multi-dimensional "satisficing" concept posited by Simon (1955) and extended on notably by Reinhard Selten (1998; 1999). However, while these procedures (and a vast number of other heuristics surveyed in Gigerenzer and Selten (1999) and Simon (1978a)) are computationally efficient and therefore procedurally rational, it is by no means assured that they are substantively rational. Indeed, both Simon and Selten explicitly differentiated their models from substantively rational models by pointing out that they were constructed to be *alternatives* to choice as solution to an optimisation problem. The benefit of having Theorem 8 above is that it provides us with a means by which to test whether these procedurally rational mappings give rise to a substantively rational choice. However, this is the topic for another paper.

5 Conclusion

The theory of rational choice has typically been simply a matter of assuming the choice of the maximal element of a set of alternatives pre-ordered by a given rational preference relation. However, when we accept that alternatives are differentiated by their attributes, and that preferences over these characteristics influence preferences over alternatives, we can no longer simply assume a rational choice amongst alternatives exists. As I have shown, for attributes to be meaningfully considered in a substantively rational decision process we

require a sufficient degree of regularity in the strength and weakness of alternatives across a broad range of characteristics if we are not to require either a rather intricate procedure for constructing preferences or one which prohibits a utility representation of the problem. I have argued that this requirement is actually quite a stringent one in a typical economic setting where benefit lies in differentiating alternatives by focussing on performing strongly in certain characteristics where other alternatives are weak, and that guaranteeing the possibility of substantively rational choice may require procedurally irrational processes.

References

Benz, M., 2007. Economics and Psychology. CESifo Seminar Series. MIT Press, Cambridge, Massachusetts, Ch. The Relevance of Procedural Utility for Economics, pp. 199–228.

Dopfer, K., Foster, J., Potts, J., 2004. Micro-meso-macro. Journal of Evolutionary Economics 14 (3), 263–279.

Earl, P., 1990. Economics and psychology: A survey. Economic Journal 100 (402), 718–755.

Frey, B., 2008. Happiness. Munich Lecture in Economics. MIT Press, Cambridge, Massachusetts.

Gigerenzer, G., Selten, R. (Eds.), 1999. Bounded Rationality. Dahlem Workshop Reports. MIT Press, Cambridge, Massachusetts.

Ironmonger, D., 1972. New Commodities and Consumer Behaviour. Cambridge University Press, Cambridge.

Jehle, A., Reny, P., 2011. Advanced Microeconomic Theory. Prentice Hall/Financial Times, London.

Kandel, E., Schwartz, J., Jessell, T., Siegelbaum, S., Hudspeth, A. (Eds.), 2013. Principles of Neural Science, 5th Edition. McGraw-Hill, New York.

Kapeller, J., Schutz, B., Steinerberger, S., 2012. The impossibility of rational conusmer choice: A problem and its solution. Journal of Evolutionary Economics 23 (1), 39–60.

Karni, E., Viero, M.-L., 2013. "Reverse Bayesianism": A choice-based theory of growing awareness. American Economic Review 103 (7), 2790–2810.

Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157.

Markey-Towler, B., 2014. The impossibility of rational consumer choice: A clarifying note, School of Economics, University of Queensland, Discussion Papers Series.

Mas-Collel, A., Winston, M., Green, J., 1995. Microeconomic Theory. Oxford University Press, Oxford.

Maslow, A., 1943. A theory of human motivation. Psychological Review 50 (4), 370–396.

Mill, J., 1863. Utilitarianism. Hackett Publishing Co, Inc, Indianopolis.

Rabin, M., 2013a. An approach to incorporating psychology into economics. American Economic Review 103 (3), 617–622.

Rabin, M., 2013b. Incorporating limited rationality into economics. Journal of Economic Literature 51 (2), 528–543.

Rubinstein, A., 2006. Lecture Notes in Microeconomic Theory. Princeton University Press, Princeton.

Selten, R., 1998. Aspiration adaptation theory. Journal of Mathematical Psychology 42, 191–214.

Selten, R., 1999. Bounded Rationality. Dahlem Workshop Reports. MIT Press, Cambridge, Massachusetts, Ch. What is Bounded Rationality?, pp. 13–36.

Simon, H., 1955. A behavioural model of rational choice. Quarterly Journal of Economics 69 (1), 99–118.

Simon, H., 1976. Method and Appraisal in Economics. Cambridge University Press, Cambridge, Ch. From Substantive to Procedural Rationality, pp. 129–148.

Simon, H., 1978a. On how to decide what to do. Bell Journal of Economics 9 (2), 494–507.

Simon, H., 1978b. Rationality as a process and as product of thought. American Economic Review 68 (2), 1–16.

Stigler, G., Becker, G., 1977. De gustibus non est disputandum. American Economic Review 67 (2), 76–90.

Strotz, R., 1957. The empirical implications of a utility tree. Econometrica 25 (2), 269–280.