Erkenn
DOI 10.1007/s10670-013-9549-x

ORIGINAL ARTICLE

Mr. Fit, Mr. Simplicity and Mr. Scope: From Social Choice to Theory Choice

Michael Morreau

Received: 2 October 2013 / Accepted: 2 October 2013

Abstract An analogue of Arrow's theorem has been thought to limit the possibilities for multi-criterial theory choice. Here, an example drawn from Toy Science, a model of theories and choice criteria, suggests that it does not. Arrow's assumption that domains are unrestricted is inappropriate in connection with theory choice in Toy Science. There are, however, variants of Arrow's theorem that do not require an unrestricted domain. They require instead that domains are, in a technical sense, 'rich'. Since there are rich domains in Toy Science, such theorems do constrain theory choice to some extent—certainly in the model and perhaps also in real science.

1 Introduction

There is an analogy between voting and multi-criterial choice. Much as voters rank candidates in an election, alternatives of one kind or another can be ordered by several choice criteria. And much as we might hope to combine voter preferences into an aggregate or 'social' ordering of the candidates, we might hope to assimilate various criterial orderings into a comparison of alternatives by their *overall* merit. Arrow's (1951) 'impossibility' theorem limits the possibilities for democratic social choice. It says that under what have seemed plausible assumptions, there is no way to derive a social ordering from voters' preferences. Consequences for multi-criterial evaluation have been explored by May (1954) and Hurley (1985), in connection with the intrapersonal determination of preferences, and by Arrow and Raynaud (1986) in industrial decision making.

Okasha (2011) proposes to apply Arrow's theorem to multi-criterial choice among rival scientific theories. He reinterprets it as an argument that there can be no

M. Morreau (✉)
UiT The Arctic University of Norway, Tromsø, Norway
e-mail: michael.morreau@uit.no

Published online: 07 November 2013

acceptable algorithm for choosing among theories on the basis of comparisons among them with respect to their fit to data, simplicity, scope and other criteria. Okasha frames this argument as a challenge to Kuhn's (1977a) thesis that different scientists may rationally choose in a range of different ways, there being no 'neutral' algorithm for theory choice.

Here I shall introduce Toy Science, a simple model of scientific theories and two criteria: fit to data and simplicity. Then I shall show that Arrow's theorem does not apply to multi-criterial theory choice in Toy Science. The reason is that there is too little variety among possible orderings of the theories by the criteria; the analogue of a crucial assumption of Arrow's theorem, namely that domains are unrestricted, is therefore unacceptable. Toy Science is simple model but it is not unrealistic. The suggestion is that Arrow's theorem does not apply to multi-criterial choice in real science either.

Even so, we will see, other limiting results closely related to Arrow's theorem might constrain theory choice to some extent. These results do not require the wide variety among criterial orderings of the *same* alternatives that is secured by Arrow's domain assumption. They require instead 'richness', a kind of variety that can be found when the criteria order *different* alternatives in a range of ways. I shall prove a variant of Arrow's theorem that exploits richness and give an example from Toy Science of a rich domain. This sets a firm limit to the possibilities for multi-criterial theory choice, at least within the model. Whether theory choice in real science is similarly constrained is another matter. That depends on whether the relevant richness is found there as well. Finding this out will require an investigation of real theories and choice criteria that I shall not undertake here. I do hope to illustrate, though, what it would take to decide the matter.

I proceed as follows. Section 2 makes the analogy between social and theory choice technically explicit, by adapting Arrow's framework to study theory choice. Section 3 introduces Toy Science. Section 4 reinterprets Arrow's theorem in theory choice. Section 5 argues that the reinterpreted theorem has no grip on theory choice because the analogue of Arrow's unrestricted-domain assumption is false. Section 6 considers the idea that Arrow's theorem might be brought to bear on theory choice by thinking of Arrow's alternatives not as rival theories but instead as 'labels' that attach to different theories, perhaps on different occasions on which there is a choice to be made. It emerges that variety among orderings of labels amounts to what is known as 'richness' of a domain. Section 7 develops a precise notion of richness, and states a variant of Arrow's theorem assuming domain richness instead of unrestrictedness (the proof is in an "Appendix"). The final Section 8 gives an example of a rich domain in Toy Science.

2 Theory Choice and Social Choice

Following Okasha (2011), this section adapts the framework of Arrow (1951) to the study of multi-criterial theory choice. We assume some set A of theoretical alternatives. These might be scientific theories, hypotheses, models or what have you. And we assume there are some criteria by which to evaluate them. Our main

question is this: can we derive from evaluations by the various criteria an ordering of the alternatives by their *overall* merit?

We will assume that an overall ordering is to be derived, if at all, from mere *comparisons* by the relevant choice criteria. That is, only ordinal information about how the alternatives measure up will factor in. In general, of course, this is not realistic. Fit to data, simplicity and other theoretical criteria can often be measured on a cardinal scale and, as Okasha points out, there are theory choice algorithms that make use of cardinal information about how theories measure up by them. One example he mentions is the *Akaike Information Criterion*, or AIC, which can be used to choose among statistical models on the basis of their fit to data and simplicity.[1]

Still, it is interesting to consider the ordinal case. For one thing, it is not difficult to see how we might in practice lack precise cardinal information, even if the relevant criteria are such, in principle, as to allow cardinal measurement. Vagueness and incompleteness of the theoretical alternatives themselves could have this effect. Take for example Darwin's theory of evolution, around the publication of *On the Origin of Species* in 1859. The notion of fitness is vague, allowing many precisifications. There is no account at all of the mechanisms of heredity. Without resolving such indeterminacy, by making concepts precise and filling in what is missing, there can be no saying exactly how well a theory fits available data, or how simple it is.

Though lacking precise cardinal information, of course we might still have more than ordinal information at hand. We might have *im*precise cardinal information. I shall not go any further, here, into questions concerning the measurability of criterial evaluations. Let us for the sake of the argument consider the case of theory choice on the basis of ordinal information. In connection with Okasha's challenge to Kuhn, anyway, this makes good sense. Kuhn explicitly argued that in historically significant cases, ordinal information is all we have to go by (Kuhn 1970, p. 147).

The criterial orderings, we will assume, are *weak orderings*. That is, they are reflexive, transitive and complete relations. A *theoretical profile* is a list of weak orderings of the set A of alternatives, one for each criterion. A profile, intuitively speaking, is one way in which all the alternatives might measure up by all the criteria. A *domain* is a set of theoretical profiles. A *theory-choice rule* is a function mapping each profile in its domain onto a weak ordering of A, intuitively an ordering of the alternatives by their overall merit.

That criterial orderings are complete might seem an unrealistic assumption. To take an example, there is an important sense in which Copernican astronomy is simpler than Ptolemaic astronomy. Its greater simplicity has to do with the number of circles in Copernican and Ptolemaic models of the solar system that—with respect to data about the angular positions of a single celestial body—are empirically equivalent. But we cannot compare the Copernican theory in this same way with, say, the phlogiston theory of combustion, or Cartesian mind–body dualism or scholastic physics. It is not simpler than they are, in the relevant sense of simplicity. Nor is less simple than they are, nor is it equally simple. There is no

[1] See Forster and Sober (1994).

comparing it with them, by this simplicity criterion, because they do not say that anything revolves on circles. If the completeness assumption is to be at all realistic then the alternatives had better be a smallish set of actual rivals, between which there is a choice to be made. Then, perhaps, we can hope to have criteria that are suitable for evaluating all of them, and complete criterial orderings.

Section 4 introduces some constraints that theory-choice rules might be expected to satisfy. Then it states Okasha's analogue of Arrow's theorem, which says that no theory-choice rule can satisfy them all. First, though, let us turn to Toy Science, the model of scientific theories and criteria. It will provide us with examples throughout.

3 Toy Science

This section sets up a simple model of scientific theories and two choice criteria: fit to data and simplicity.

3.1 Theories

We will assume a space of possible states of affairs, or *possible worlds*. Scientific theories distinguish possibilities with which they are compatible from those which they exclude. Idealizing away vagueness and other kinds of indeterminacy, we will suppose that this distinction is completely sharp, so that theories determine *sets* of compatible worlds. In Toy Science, we simply identify theories with these *propositions*, or sets of possible worlds. We will consider only finite propositions.

3.2 Fit to Data

Worlds, we assume, are to a greater or lesser extent similar to one another. Let d be a Euclidean metric on the worlds. For any world u, $d(u,u) = 0$, while for any other world v, $d(u, v) > 0$. Intuitively, $d(u, v)$ is the degree to which u and v differ. The following example will run throughout:

Example 1 Assume there is within the space of worlds an equilateral triangle whose vertices are the possible worlds u, v, and w. Along each edge there are other worlds. Passing from u to v, we encounter before the halfway point one such world, call it uv. With elementary geometry, some simple facts can be established that will be used later on: $d(u, uv) < d(v, uv) < d(w, uv)$. After the halfway point there is another world, vu; and similarly spaced along the other edges of the triangle are other worlds: vw, wv, uw, and wu. Similar facts concerning them will also be used.

Existing observations and the results of experiments we will lump together as *bodies of data*. These, like theories, are finite propositions. A world is fully compatible with a proposition if it is among the worlds of that proposition, but less compatible the more distant it is, in the sense of d, from them. More generally, let the degree of incompatibility of a theory T with a body of data D be the least distance between their worlds:

Definition 2 (*incompatibility of theory with data*)

$$in(T, D) = min\{d(v, w): v \in T, w \in D\}$$

A theory T is fully compatible with data D, intuitively speaking, if T and D can both be correct. That is, some world is fully compatible with both of them: $T \cap D \neq \{\}$. Then, and only then, is the degree $in(T, D)$ of their incompatibility O. The greater $in(T, D)$ is, the more dissimilar do T and D require the world to be.[2]

Now, according to the following notion of comparative fit, those theories fit the data better that are less incompatible with it:

Definition 3 (*comparative fit*)

$$S \,{}_f\!\geq T \quad \text{if} \quad in(S, D) \leq in(T, D).$$

Notice that D is left implicit in the notation ${}_f\!\geq$. In what follows, comparisons of fit will be made in relation to some contextually determinate body of data. To illustrate this notion of comparative fit, I shall now develop Example 1.

Example 4 By Definition 2, the theories $\{u\}$ and $\{v\}$ are both fully compatible with $D = \{u, v\}$, but $\{w\}$ is not. That is: $in(\{u\}, D) = in(\{v\}, D) = 0 < in(\{w\}, D)$. By Definition 3, we have the following ordering of these theories $\{u\}$,$\{v\}$, and $\{w\}$ by their fit to D: $\{u\} \,{}_f\!\approx \{v\} \,{}_f\!> \{w\}$.[3] With other bodies of data we have other orderings. Relative to $D = \{uv\}$, for example, the ordering is: $\{u\} \,{}_f\!> \{v\} \,{}_f\!> \{w\}$. The following thirteen bodies of data produce all thirteen weak orderings of the three theories $\{u\}$,$\{v\}$, and $\{w\}$ by their fit to the data: $\{u, v, w\}$, $\{u, v\}$, $\{v, w\}$, $\{u, w\}$, $\{uv\}$, $\{vu\}$, $\{vw\}$, $\{wv\}$, $\{uw\}$, $\{wu\}$, $\{u\}$, $\{v\}$, and $\{w\}$.

3.3 Simplicity

This section develops a way of comparing theories by how complex they require the world to be. Say any given world w has a degree of complexity: $c(w) \geq 0$. A theory, we will say, is as complex as it requires the world to be: the complexity of T is that of the least-complex world compatible with T: $c(T) = min\{c(w): w \in T\}$. Now we count those theories as simpler that require less complexity:

Definition 5 (*comparative simplicity*) $S \,{}_s\!\geq T$ if $c(S) \leq c(T)$.

For an intuitive example, suppose a world with both immaterial souls and material substances is more complex than a world with just material substances. Then, by this reckoning, a theory saying there are both immaterial souls and material substances is less simple than another theory that, while perhaps not excluding the possibility of souls, does not *require* that there are any. The following rather trivial

[2] John Bigelow (1977, p. 462) found this same notion useful in accounting for probability within a possible-worlds framework.

[3] As usual, $\approx$ means that both $\geq$ and $\leq$; and $>$ means that $\geq$ but not $\leq$.

kind of case will be important in Section 8, when we construct a rich domain in Toy Science:

Example 6 From now on, assume that u, v, and w are such that $c(u) = c(v) < c(w)$. (Say, u and v contain just material substances, while w contains both material substances and immaterial souls.) The simplicity ordering of the theories $\{u\}$,$\{v\}$, and $\{w\}$ is: $\{u\}\ {}_s\approx\ \{v\}\ {}_s> \{w\}$.

Now we have some examples of theoretical profiles in Toy Science, and of domains for theory choice:

Example 7 Let A contain the seven non-empty subsets of $\{u, v, w\}$. A theoretical profile is an ordered pair of weak orderings of these theories. The first component is an ordering by their fit to some body D of data. For example, with $D = \{v\}$, the ordering ${}_f\geq$ by fit is:

$$\{u,v,w\}_f \approx \{u,v\}_f \approx \{v,w\}_f \approx \{v\}$$
$$_f >$$
$$\{u,w\}_f \approx \{u\}_f \approx \{w\}.$$

The simplicity ordering ${}_s\geq$ is:

$$\{u,v,w\}_s \approx \{u,v\}_s \approx \{v,w\}_s \approx \{u,w\}_s \approx \{u\}_s \approx \{v\}$$
$$_s >$$
$$\{w\}.$$

Together, these two orderings make up a profile $[{}_f\geq, {}_s\geq]$. Other data sets determine other orderings of A by fit to the data. Each such ordering, together with the above simplicity ordering, makes up a theoretical profile. A domain for theory choice is a set of profiles, one for each data set. Notice that all profiles in a domain share the same simplicity ordering.

Our question is now: are there satisfactory theory-choice rules for these domains? That is, are there functions that map the profiles in their domains onto weak orderings of A, and meet expectations we may reasonably have of them? In the next section, we will have an argument according to which there are no satisfactory theory-choice rules.

4 Arrovian Nihilism About Theory Choice

Arrow's (1951) theorem tells us that, under certain assumptions, there is no combining various individual preferences among some alternatives into a single 'social' ordering. Okasha (2011) reinterprets this theorem as an argument that there is no acceptable way to order rival scientific theories by their overall merit on the basis of comparisons among them with respect to theoretical values including fit to

data and simplicity. This section states Okasha's reinterpretation of Arrow's theorem.

A first assumption of Okasha's is that the overall ordering of any given pair of alternatives depends only on how they measure up by the various criteria. That is, there is to be no change in the overall ordering of this pair without a change in how they compare with respect to one another by some or other criterion. A profile, remember, is a list of orderings of all of the alternatives, one ordering for each criterion. The relevant assumption is:

Independence of irrelevant alternatives Whenever two profiles completely agree, criterion for criterion, as far as any given pair of alternatives is concerned, the comparative overall merit of this pair must also be the same, relative to these profiles.

Theory-choice rules, although they are functions, need not be consistent in their treatment of a pair of alternatives as we move from one profile to the next within the domain. The overall ordering of a pair of alternatives might in principle depend not only on how they compare by the various criteria to one another, but also on how they compare to 'irrelevant' alternatives outside of the pair. This is what Independence rules out.

Let us say that a choice criterion is a *dictator* if it is completely decisive in the sense that an alternative always counts as strictly better overall, when it is strictly better by this one criterion. A further assumption of the theorem is:

Non-dictatorship No criterion is a dictator.

Acceptable theory-choice rules, according to this assumption, are those that balance the various merits and demerits of the alternatives against each other.

You might on empiricist grounds have thought that fit to data is decisive—if not when choosing a theory for some applied purpose, then certainly in science, when we are after the truth. But fit does not dictate overall merit even then. Theories can be brought into agreement with observations and experimental results by adjusting parameters. Scientific data are generally noisy, though. So if we single-mindedly pursue fit to available data, without regard to other criteria, we will end up preferring overly complicated theories, whose many parameters we can tune to fit the noise. These theories might agree with every error in the data; but they will fit underlying facts and future data less well than do other, simpler theories. They will *overfit* the data. To avoid this, acceptable choice rules must balance fit against simplicity. And they must do so not in spite of the special importance of accuracy in science but precisely because of it, because balance is what secures accuracy in the long run.[4] Accuracy, then, doesn't dictate overall merit. Plainly, simplicity doesn't either; and nor do scope or any other of Kuhn's (1977a) 'objective' criteria of consistency, fruitfulness and so on.

Another assumption is:

[4] I take this point from Forster and Sober (1994).

Weak Pareto If one theory is strictly better than another by every criterion, then it is strictly better overall.

This requires theory-choice rules to respect unanimity among the criteria.

The remaining assumption concerns the variety to be found among profiles within the domains of theory-choice rules. I shall state this assumption more precisely than I have done the others, since we will consider it closely in the next section:

Unrestricted domain For any weak orderings ${}_f\geq$, ${}_s\geq$… of the alternatives A, the domain of the theory-choice rule includes the profile $[{}_f\geq, {}_s\geq, \ldots]$.

This requires theory-choice rules to accept as input a wide variety of profiles.

Arrow's theorem states that if A contains at least three alternatives, and there are finitely many criteria, then one of the four assumptions has to be false.[5] Okasha argues that all four are requirements for the acceptability of theory-choice rules and that Arrow's theorem poses a threat to the rationality of science.

5 Domain Restrictions in Theory Choice

In this section, I shall show that it is absurd to impose unrestricted domains on theory-choice rules for Toy Science. The variety among profiles is nowhere near what is needed for Arrow's theorem to get a grip. I argue in a forthcoming article that Arrow's theorem does not apply to real science either, and the reason is the same. The analogue of the domain assumption is inappropriate.

Notice, first, that unrestricted domains are really big. Suppose A contains seven alternatives—say, the non-empty subsets of $\{u, v, w\}$ of Example 7. That is not a large number when choosing, say, among statistical models. There are 47,293 weak orderings of seven items.[6] Even with just the two criteria of fit and simplicity, as in Toy Science, an unrestricted domain for A includes $472{,}93^2$, or 2,236,627,849 profiles. That is almost 2¼ billion. We might like to take into account an additional criterion, say scope. With three criteria there are over a 100 trillion profiles. Kuhn (1977a) thought that five 'objective' theoretical values form the basis for theory choice: accuracy, consistency, scope, simplicity, and fruitfulness; but he thought other criteria are sometimes relevant as well. Suppose we evaluate the seven theories in A using seven criteria. Then Unrestricted domain requires an acceptable rule for choosing among these theories to reckon with a cool *five hundred million trillion trillion* profiles.

Okasha finds Unrestricted domain reasonable enough:

> [It] seems unexceptionable—however the theories are ranked by the various criteria, the rule must be able to yield an overall ranking. There should be no apriori restriction on the permissible rankings that are fed into the rule. (Okasha 2011, p. 92)

[5] For a precise statement of the assumptions of Arrow's theorem and a proof, see any standard text such as Gaertner (2009).

[6] See Bailey (1998).

Considering particular theories along with criteria suitable for evaluating them, though, it seems that there should be severe apriori restrictions. Toy Science illustrates:

Example 8 Let A contain the seven non-empty subsets of $\{u, v, w\}$, as in Example 7. And let the possible data sets be all non-empty subsets of the nine worlds in the triangle: u, v, w, uv, vu, vw, wv, uw, wu. Domains contain profiles determined by some or all of these data sets.

With these theories, criteria and data sets, there are at most 511 possible profiles (corresponding to the $2^9 - 1$ or 511 data sets, the nonempty sets of worlds). There is nothing even remotely like the 2¼ billion of an unrestricted domain.

The domain restrictions in Toy Science are not artificial, nor are they any sort of imposition from outside. Rather, they emerge naturally from the theoretical alternatives and choice criteria themselves. It is instructive to see how. Let us say that one theory S is *logically stronger* than other, T, if $S \subseteq T$. The stronger theory says more about what the world is like. Consider two theories, neither of which is logically stronger than the other—say, $\{u, v\}$ and $\{v, w\}$. The first theory says that the actual world is either u or v, and the second that it is either v or w. It is not difficult to see that, depending on the data, either theory can fit the data better than does the other (take as data sets the set-theoretic differences between the two theories, in this case $\{u\}$ and $\{w\}$). With other theories, though, there is not the same variety among possible orderings by fit to data. If S is logically stronger than T, it is easily seen that, no matter what the data are, $T \underset{f}{\geq} S$. Profiles had better rank two such theories accordingly. What sense can there be in admitting for aggregation criterial orderings that get things *impossibly* wrong?

That some theoretical alternatives are logically stronger than others is not just an artifact of the way theories are rendered in Toy Science as sets of possible worlds. Such logical relations among theoretical alternatives, and corresponding restrictions on orderings by fit to data, are also found in examples from real science. In statistical modeling, for example, the model LIN, the class of all linear curves, is nested within the parabolic model PAR, because each linear curve is also a parabolic curve (with one adjustable parameter set to 0). For this reason, LIN—or, if you will, the hypothesis that the best curve is linear, some curve within this model—is logically stronger than PAR. In fact, nesting is common in normal science. It arises when we elaborate models so as to accommodate data.[7]

Another domain restriction in Toy Science is inherent in the fact that which theories are simpler than which, in the relevant ontological sense of 'simplicity', does not depend on anything else, in the way that which theories fit the data better depends on the data. It is just a matter of which theories they are. Worlds contain whichever kinds of things they contain. Theories, sets of worlds, have whichever elements they have. Whether one theory is simpler than another is completely determined by this, so it is inherent in the theories. There is but one correct way to order them by their simplicity.

[7] Forster (2004) discusses examples of nested models in science. For more examples of domain restrictions in real science, see Morreau forthcoming.

Additional domain restrictions can easily be imagined that are not inherent in the theories and the criteria used to evaluate them. For example, a restriction on admissible data will tend to restrict variety among possible orderings by fit to the data. Be this as it may, from the inherent restrictions it is plain that, as far as the alternatives and criteria of Toy Science are concerned, the idea of imposing an unrestricted domain on theory-choice rules is absurd.

Arrow's domain assumption is known to be unnecessarily strong. Many domains, though restricted, retain enough variety to obtain a similar result. Domains for theory choice do not seem to, though. Variants of Arrow's theorem that tighten up its domain assumption typically still require the domain to be unrestricted at least with respect to some triple of alternatives.[8] If there is any criterion that can order the alternatives in just one way that is out of the question. There is no such triple. Simplicity is such a criterion, in the example.

In conclusion, Arrow's theorem does not limit the possibilities for multi-criterial theory choice in Toy Science. There isn't enough variety in the criterial orderings for the theorem to get a grip. Lacking any examples at all of theories and criteria that do generate enough variety, whether Toy examples or real ones, you have to wonder whether Arrow's theorem has any relevance at all for multi-criterial theory choice.

6 Representations?

Suppose we think of the elements of A not as themselves theories, but rather as names or 'labels' that attach to different theories—perhaps on different occasions on which there is a choice to be made. A profile is then a list of orderings of these labels, one for each choice criterion. Say a profile of labels is *admissible* if, among the theories to which the labels may be attached, there are some that really are ordered, by the criteria, in the way that the profile specifies.[9] Then there might be variety among admissible profiles of labels even if the criterial orderings of any given theories are severely restricted. It derives from variety among the different theories to which the labels may be attached. There might even be enough it for an Arrow-style impossibility result.

This *representational* interpretation of Arrow's framework is not standard. Standardly, the social alternatives are not names or labels. They are candidates in elections, public projects that might be carried out, distributions of income and labour requirements among members of society, and so on. But some think it correct to think of Arrow's alternatives as labels, and that might explain their acquiescence in the idea that domains for theory choice are unrestricted.[10] Getting clear about the

[8] The *chain property* is an example. See Campbell and Kelly (2002, p. 41).

[9] I leave intuitive, for now, the idea of ordering some alternatives in some particular 'way'. It will become precise in the next section, with the notions of *patterns* and their *realization* by profiles.

[10] I do not know of any explicit discussion in the literature of the representational interpretation, but it does seem to be a part of the folklore. An anonymous reviewer of Morreau forthcoming, rejecting the argument of the previous section that inherent domain restrictions make Arrow's theorem inapplicable to theory choice, wrote that we 'must' understand Arrow's alternatives as names or labels, even though

difference between a representational interpretation and the standard one might promote an appreciation of the difference between Arrow's theorem and the close relative that I shall discuss in the next section, before demonstrating its relevance to theory choice in Section 8. In the remainder of this section, then, I shall briefly consider the consequences of thinking of Arrow's alternatives as labels.

Unrestricted domain, on a representational interpretation, seems to be a weaker assumption than it standardly is. Standardly, in connection with theory choice, it requires that some given set of alternatives can be ordered in every 'logically possible' way by the relevant criteria. Every weak ordering of these alternatives is a real possibility. On a representational interpretation, the requirement seems to be just that for each logically possible way of ordering alternatives there are, among the alternatives to which the labels may be attached, some that the criteria do in fact order that way. The crucial difference is that it doesn't have to be the same ones in each case. A related kind of requirement has been studied on a standard interpretation as domain 'richness'.

Independence, on the other hand, on a representational interpretation, seems to be a stronger assumption than it standardly is. Standardly interpreted, in connection with theory choice, it requires that the overall comparison among any given pair of alternatives just depends on how these two alternatives measure up by the various criteria. The terms of the dependence might be different for different pairs, though; for example, the criteria might carry different weights when used to evaluate one pair than when used to evaluate another pair. On a representational interpretation, Independence will rule out this variation from pair to pair because it requires consistency among all pairs to which the labels may be attached. This seems to be what is standardly called 'neutrality'.

A representational rejigging Arrow's framework is, I think, both unwise and unhelpful. It invites confusion among notions that have long been distinguished and studied on a standard interpretation. Better to leave the framework as it is and have separate notions of domain richness and neutrality. That is what I shall do here. The next section introduces suitable notions and implicates them in an impossibility theorem closely related to Arrow's. Section 8 finds a rich domain in Toy Science.

7 Impossibility with a Restricted Domain

It will help to have a precise notion of a 'way' of ordering alternatives. Let X, Y, Z… be (logical) variables that can stand in for alternatives in A. (Think of them, if you like, as the 'labels' of the previous section.)

Definition 9 A *pattern* is a list of weak orderings of some set of variables.

Footnote 10 continued
Arrow does not seem to have done so. Perhaps it is easy to fall into a representational interpretation of Arrow's alternatives if one thinks of choice rules computationally, as procedures to be applied to different sets of alternatives on different occasions, in much the same way that a voting procedure is used year after year with different slates of candidates.

Technically, a pattern is just like a profile except that it orders variables instead of alternatives. A pattern is *suitable* for a domain if it has just as many orderings as the profiles of this domain have choice criteria. For example, with just the two criteria of Toy Science, theoretical profiles are pairs of orderings of theories, domains are sets of these, and the suitable patterns are pairs of orderings of variables. One such pattern is for instance $[Y > X \approx Z, X \approx Y > Z]$, a pair of weak orderings of the set $\{X, Y, Z\}$ of variables.

Next, we need a precise notion of what it is for some theories to be ordered by the relevant criteria in the way specified by a pattern.

Definition 10 Theoretical profile P^* *realizes* pattern P if there is an embedding of P into P^*.

By an *embedding* I mean a structure-preserving mapping that assigns to each of the variables of the pattern a unique alternative in A. Intuitively speaking, a profile realizes a pattern if it orders some or other theories just as the pattern specifies.

Example 11 Let P^* be the theoretical profile, corresponding to data $D = \{v\}$, from Example 7. We have, in P^*, $\{v\} \ {}_f\!> \{u\} \ {}_f\!\approx \{w\}$, and $\{u\} \ {}_s\!\approx \{v\} \ {}_s\!> \{w\}$.

Mapping the variable X onto $\{u\}$, Y onto $\{v\}$, and Z onto $\{w\}$, P^* can be seen to realize the pattern $[Y > X \approx Z, X \approx Y > Z]$. We have seen in Example 4 that with suitable choice of D we can obtain any weak ordering of the theories $\{u\}$,$\{v\}$, and $\{w\}$ by their fit to D. Therefore every pattern of the form $[\ldots, X \approx Y > Z]$ (where … is a weak ordering of X, Y and Z) is realized by some or other profile of the sort discussed in Example 7.

I have introduced the notion of a pattern quite generally, for any sets of variables. The following notion of domain richness concerns patterns of *triples* of variables:

Definition 12 A domain is *rich* if for every suitable pattern P of three variables, there is some profile in this domain that realizes P.

In the next section, I shall develop the running example to give an example of a rich domain for theory choice. Meanwhile, let us in the remainder of this section turn to the notion of neutrality. Independence, as we have seen, requires consistency of aggregation for pairs of alternatives: whenever two profiles order any given pair in exactly the same way, the aggregate ranking of this pair has to be the same as well, relative to these profiles. Neutrality requires this, but also that aggregation procedures are insensitive to the *identities* of alternatives:

Definition 13 A theory-choice procedure is *(strongly) neutral* if whenever some alternatives x and y are ordered in one profile, criterion for criterion, just as some (perhaps different) alternatives z and w are ordered in some (perhaps different) profile, the comparative overall merit of x and y relative to the first profile is the same as that of z and w relative to the second profile.[11]

Social decision-making lacks neutrality if voters' preferences determine social orderings differently in different cases—if say a simple majority is allowed to

[11] For a precise statement, see a standard text such as Gaertner (2009), or see Morreau forthcoming.

overturn the status quo in an ordinary case while a supermajority is required in an extraordinary one. Our evaluation of scientific theories lacks neutrality if the various criteria can go together differently depending on which theories we are choosing among. There needn't be anything unscientific about this. There is no reason why the relative importance of fit and simplicity should be the same, say, in ecology as it is in physics. Considering sufficiently similar cases though, within some scientific field or sub-field, we might expect choices among theories to be made in the same way, just as we might expect social decisions to be made in the same way in all routine cases, and the same way in all extraordinary ones. Suitably restricted, neutrality does seem a part of what it is for a choice procedure to be principled.

Now it turns out that with rich domains there is a limit, closely related to that revealed by Arrow's theorem, to the possibilities for neutral choice procedures.

Fact 14 Suppose the set of theoretical criteria is finite. No theory-choice rule with a rich domain satisfies neutrality, weak Pareto and non-dictatorship.[12]

There is a proof of this fact in the "Appendix".

Are theoretical domains ever so rich as to preclude neutral choice procedures? In the next section, we will see that sometimes, in Toy Science anyway, they are.

8 A Rich Domain in Toy Science

This section constructs a rich domain for theory choice by further developing the running example. The triangle of worlds from the earlier examples we now name Δ_2; its worlds: u, v,…, uv,… we rename accordingly: u_2, v_2,…, uv_2,…. We assume that there are, in our space of possible worlds, three other such equilateral triangles Δ_1, Δ_3 and Δ_4. The difference between Δ_2 and these other triangles is in the comparative simplicity of their worlds. In Δ_2 we have: $c(u_2) = c(v_2) < c(w_2)$. Δ_1, Δ_3, and Δ_4 are to be such that:

$$\text{in } \Delta_1 : \; c(u_1) = c(v_1) = c(w_1);$$
$$\text{in } \Delta_3 : \; c(u_3) < c(v_3) = c(w_3); \quad \text{and:}$$
$$\text{in } \Delta_4 : \; c(u_4) < c(v_4) < c(w_4).$$

Now I shall construct a rich domain for theory choice using these triangles:

Example 15 Let A consist of the twelve theories $\{u_i\}$, $\{v_i\}$, and $\{w_i\}$, $1 \leq i \leq 4$. Let the domain $\Re$ contain all theoretical profiles of A determined by each of the fifty-two (=13 × 4) data sets, by analogy with the thirteen data sets in Example 4.

[12] Fact 14 is a variant of the 'single profile' impossibility theorems pioneered by Parks (1976) and (independently) by Kemp and Ng (1976). These respond to an objection leveled against Arrow's 'multi-profile' framework, according to which there is only a single profile of individual preferences that needs to be taken into account in social choice, namely the preferences that the members of society happen actually to have. In theory choice, as we have seen, there might be a certain amount of variety among the profiles that it makes sense to submit to a theory-choice rule, though it falls well short of what is required by Arrow's domain assumption.

$\Re$ is rich. To be shown is that each suitable pattern of three variables is realized by some profile in $\Re$. Consider any such pattern P. Up to alphabetic variation, there are just four weak orderings of three variables. They are: (1) $X \approx Y \approx Z$, (2) $X \approx Y > Z$, (3) $X > Y \approx Z$, and (4) $X > Y > Z$. We may assume without loss of generality that the second ordering of P is one of these: (i). Mapping X onto $\{u_i\}$, Y onto $\{v_i\}$, and Z onto $\{w_i\}$, with suitable choice of D we can obtain any weak ordering of these three theories by their fit to D. One such D determines a profile that realizes P. For example, suppose $P = [Z > X \approx Y, X > Y \approx Z]$. Its second component is ordering (3), so we map X onto $\{u_3\}$, Y onto $\{v_3\}$, and Z onto $\{w_3\}$. Letting $D = \{w_3\}$, we have: $\{w_3\}\ {}_f\!> \{u_3\}\ {}_f\!\approx \{v_3\}$. Since $\{u_3\}\ {}_s\!> \{v_3\}\ {}_s\!\approx \{w_3\}$, P is realized by the profile determined by D.

We have seen that $\Re$ is a rich domain. Fact 14 tells us that, under what appear to be minimal further assumptions, there is no neutral theory-choice algorithm for $\Re$. Here, then, is a limit to the possibilities for theory choice in Toy Science. Whether real theory-choice procedures run up against this limit has yet to be seen. This section has illustrated what it would take to show that they do.

Acknowledgments I thank for helpful comments Malcolm Forster, Aidan Lyon, Samir Okasha, John Weymark, and an anonymous reviewer.

Appendix

Proof of Fact 14 Suppose the theoretical criteria are finite in number and that some theory-choice rule f has a rich domain and satisfies neutrality and weak Pareto. To be shown is that f has a dictator.

First, letting the criteria be 1,…,*n*, because the domain is rich we can find a series of profiles $R_0,\ldots, R_n$ and pairs s_j, $t_j \in A$ such that for each $0 \leq j \leq n$:

$$t_j <_i^j s_j \quad \text{if} \quad 1 \leq i \leq j, \quad \text{and} \quad s_j <_i^j t_j \quad \text{if} \quad j < i \leq n.$$

Here, $\leq_i^j$ is the ith criterial ordering of profile R_j. Writing $\leq^j$ for $f(R_j)$, by weak Pareto we have $s_0 <^0 t_0$ and $t_n <^n s_n$. Counting up from 0 to *n*, let d be the first j such that not: $s_j <^j t_j$. By choice of d and completeness of $\leq^d$:

$$s_{d-1} <^{d-1} t_{d-1}, \quad \text{and} \tag{1a}$$

$$t_d \leq^d s_d \tag{1b}$$

This criterion d is a dictator. Consider any profile P in the domain and any alternatives a, $c \in A$ such that $a <_d^P c$ (which is to say that, according to P, c is strictly better than a by criterion d). To be shown is that:

$$a <^P c \tag{2}$$

(in the sense of the overall ordering that f assigns to P, c is strictly better than a). Since the domain is rich, there is a profile Q and there are alternatives α, β, $\gamma \in A$, such that for all criteria i:

$$a \leq_{\mathrm{i}}^{P} c \quad \text{iff} \quad \alpha \leq_{\mathrm{i}}^{Q} \gamma; \quad \text{and} \quad c \leq_{\mathrm{i}}^{P} a \quad \text{iff} \quad \gamma \leq_{\mathrm{i}}^{Q} \alpha; \tag{3}$$

$$\alpha <_{\mathrm{i}}^{Q} \beta \quad \text{and} \quad \gamma <_{\mathrm{i}}^{Q} \beta, \quad \text{if} \quad 1 \leq \mathrm{i} < \mathrm{d}; \tag{4}$$

$$\alpha <_{\mathrm{d}}^{Q} \beta <_{\mathrm{d}}^{Q} \gamma; \quad \text{and}$$

$$\beta <_{\mathrm{i}}^{Q} \alpha \quad \text{and} \quad \beta <_{\mathrm{i}}^{Q} \gamma, \quad \text{if} \quad \mathrm{d} < \mathrm{i} \leq \mathrm{n}.$$

By (3) and neutrality, it is sufficient for (2) that $\alpha <^{Q} \gamma$, and for this it is by transitivity of $\leq^{Q}$ sufficient that:

$$\alpha \leq^{Q} \beta, \quad \text{and:} \tag{5}$$

$$\beta <^{Q} \gamma. \tag{6}$$

By inspection of R_{d} and (4), for all criteria i:

$$\alpha \leq_{\mathrm{i}}^{Q} \beta \quad \text{iff} \quad \mathrm{t_d} \leq_{\mathrm{i}}^{\mathrm{d}} \mathrm{s_d}, \quad \text{and} \quad \beta \leq_{\mathrm{i}}^{Q} \alpha \quad \text{iff} \quad \mathrm{s_d} \leq_{\mathrm{i}}^{\mathrm{d}} \mathrm{t_d}.$$

Now (5) follows by neutrality from (1b). Also, by inspection of $R_{\mathrm{d}-1}$ and (4), for all i:

$$\beta \leq_{\mathrm{i}}^{Q} \gamma \quad \text{iff} \quad \mathrm{s_{d-1}} \leq_{\mathrm{i}}^{\mathrm{d}-1} \mathrm{t_{d-1}} \quad \text{and} \quad \gamma \leq_{\mathrm{i}}^{Q} \beta \quad \text{iff} \quad \mathrm{t_{d-1}} \leq_{\mathrm{i}}^{\mathrm{d}-1} \mathrm{s_{d-1}}.$$

Now (6) follows by neutrality from (1a).

This completes the demonstration that (2). We have seen that d is a dictator.[13]

References

Arrow, K. (1951). *Social choice and individual values*. New York: Wiley.

Arrow, K., & Raynaud, H. (1986). *Social choice and multicriterion decision-making*. Cambridge Ma. and London: The MIT Press.

Bailey, R. W. (1998). The number of weak orderings of a finite set. *Social Choice and Welfare, 15*, 559–562.

Bigelow, J. C. (1977). Semantics of probability. *Synthese, 36*, 459–472.

Campbell, D. E., & Kelly, J. S. (2002). Impossibility theorems in the Arrovian framework. In K.J. Arrow, A.K. Sen & K. Suzumura (Eds.), (pp. 35–94).

Forster, M. (2004). Chapter 3: Simplicity and unification in model selection. http://philosophy.wisc.edu/forster/520/Chapter%203.pdf.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science, 45*, 1–35.

Gaertner, W. (2009). *A primer in social choice theory* (revised ed.). Oxford: Oxford University Press.

Geanakoplos, J. (2005). Three brief proofs of Arrow's theorem. *Economic Theory, 26*, 211–215.

Hurley, S. (1985). Supervenience and the possibility of coherence. *Mind, 94*, 501–525.

Kemp, M. C., & Ng, Y. K. (1976). On the existence of social welfare functions, social orderings and social decision functions. *Economica, 43*, 59–66.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Kuhn, T. (1977a). Objectivity, value judgment, and theory choice. In his 1977b, (pp. 320–339).

Kuhn, T. (1977b). *The essential tension*. Chicago: University of Chicago Press.

[13] This is an adaptation and simplification of one of John Geanakoplos's (2005) proofs of Arrow's theorem.

May, K. O. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica, 22*, 1–13.

Morreau, M. Theory choice and social choice: Kuhn vindicated. *Mind* (forthcoming).

Okasha, S. (2011). Theory choice and social choice: Kuhn versus Arrow. *Mind, 120*, 83–115.

Parks, R. P. (1976). An impossibility theorem for fixed preferences: A dictatorial Bergson–Samuelson welfare function. *Review of Economic Studies, 43*, 447–450.