Okasha, S. (2015). On Arrow's Theorem and Scientific Rationality: Reply to Morreau and Stegenga. Mind, 124(493), 279-294. DOI: 10.1093/mind/fzu177

On Arrow's Theorem and Scientific Rationality: reply to Morreau and Stegenga

SAMIR OKASHA
University of Bristol
Samir.Okasha@bristol.ac.uk

Abstract

In a recent article I compared the problem of theory choice, in which scientists must choose between competing theories, with the problem of social choice, in which society must choose between competing social alternatives. I argued that the formal machinery of social choice theory can be used to shed light on the problem of theory choice in science, an argument that has been criticised by Michael Morreau and Jacob Stegenga. This article replies to Morreau's and Stegenga's criticisms.

1. Introduction

In my article 'Theory Choice and Social Choice: Kuhn versus Arrow', I explored an analogy between the problem of theory choice, in which scientists must choose between competing theories or hypotheses, and the problem of social choice, in which society must choose between competing social alternatives or 'ways society might be' (Okasha 2011). Theory choice presents a problem because there are multiple desiderata that we would like a theory to satisfy, e.g. fit-with-the-data and simplicity, which may pull in different directions, as Kuhn (1969, 1977a) famously argued. Social choice presents a problem because there are multiple individuals in society, whose preferences over the alternatives may not coincide.

By identifying social alternatives with competing theories, and individuals with desiderata (or 'criteria of theory choice'), I showed that the theory choice problem and the social choice problem have essentially the same structure. In each case the problem is one of aggregating individual rankings into an overall ranking. This aggregation problem has been extensively studied in social choice theory, which suggests using formal results from that field, such as Arrow's impossibility theorem, to study the problem of theory choice in science.

As is well-known, Kenneth Arrow (1951) argued that any reasonable aggregation procedure, or 'social choice rule' as I called it, should satisfy four conditions (**U**niversal Domain, **N**on-dictatorship, **I**ndependence of Irrelevant Alternatives and Weak **P**areto); he then proved that these conditions cannot in fact be jointly satisfied, so long as the set of social alternatives contains at least three members. In Okasha 2011, I argued that the analogues of Arrow's four conditions are reasonable requirements to impose on a theory choice rule, and thus that, *prima facie*, an impossibility theorem should hold for theory choice. This conclusion is puzzling, given that scientists do (apparently) manage to make all-things-considered judgments about which of a set of competing theories is the best; and such judgments often appear perfectly rational.

To resolve the puzzle, I considered a number of possible 'escape routes' from Arrow's impossibility result. I argued that one particular escape route, pioneered originally by Sen (1970, 1977) in relation to social choice, is applicable to at least some cases of theory choice in science. Sen's idea was to 'enrich the informational basis', by allowing as input into the social choice rule more than the merely ordinal rankings of alternatives, which permit no interpersonal comparisons, that Arrow employed. I showed how this escape route is implicit in two well-known approaches to theory choice in the philosophy of science (statistical model selection and Bayesianism), thus explaining how they avoid the threat of impossibility.

Stegenga (2014) and Morreau (2014) both offer critical assessments of my arguments for which I am grateful. They arrive at diametrically opposite conclusions. Stegenga agrees with me that Arrow's impossibility result does potentially apply to theory choice, but disagrees with my suggestion that 'enriching the informational basis' offers a potential way out. Thus he thinks that it 'remains puzzling' how rational theory choice is possible (p.??). Morreau argues, by contrast, that there is no threat of impossibility in the first place, since the analogue of Arrow's condition **U** (unrestricted domain) does not apply to theory choice. (He argues similarly in Morreau 2013). I reply to their arguments in turn.

2. Stegenga

Before turning to Stegenga's main criticisms of my paper, it is worth clearing up a number of technical errors that he has introduced into the discussion. In Part 3 of his paper, Stegenga describes in his own words Sen's 'informational enrichment' strategy for avoiding Arrovian impossibility. Sen's key move was to use profiles of utility functions of the form $<u_{1,}...,u_{n}>$, rather than profiles of preference orderings of the form $<R_{1,}...,R_{n}>$, as input into the aggregation rule, where u_{i} denotes individual *i*'s utility function over the set of social alternatives in question; the aggregation rule is then known as 'a social welfare functional'. Stegenga unhelpfully describes a utility function as an assignment of real numbers to 'choices', rather than to alternatives, but this is a minor matter.

Less minor is Stegenga's discussion of how, in Sen's framework, Arrow's original condition **I** can be decomposed into sub-components, namely independence of irrelevant utilities (**IIU**) and ordinal non-comparability of utility (**ONC**), whose conjunction is logically equivalent to Arrow's condition **I**, as I described in my 2011 article (following Sen). Stegenga's exposition of this decomposition differs from the standard one in two respects. He believes that **ONC** can be sub-divided into two further conditions (**O** and **NC**); and he confusingly refers to the first sub-component as 'irrelevance of alternatives' (**IA**), instead of **IIU**. Moreover, when he explains what **IA** says, it turns out that he is actually treating **IA** to be identical to Arrow's condition **I** itself! (He writes: '**IA** holds that how a theory choice algorithm ranks Copernican heliocentrism to Ptolemaic geocentrism should only depend on how the theoretical virtues rank Copernican heliocentrism to Ptolemaic geocentrism' (Stegenga 2014 p.??). So in Stegenga's discussion, **IA** is not a 'sub-component' of Arrow's independence condition **I** at all, but rather just **I** itself. This part of Stegenga's discussion is therefore confused.

Moreover, Stegenga does not explain how condition **ONC** can be split into two. **ONC** is an invariance requirement on the social welfare functional: it says that given two profiles of utility functions $<u_{1,}...,u_{n}>$ and $<v_{1,}...,v_{n}>$, if each v_{i} is a monotonic transformation of each u_{i}, not necessarily the same one for each *i*, then the social welfare functional must map the two profiles onto the same ranking. (If 'monotonic' were replaced with 'positive linear', the resulting condition would be cardinal non-comparability (**CNC**)). Stegenga asserts that it is possible to split **ONC** into two sub-conditions, **O** and **NC**, but does not provide precise formulations of them. I do not see that this is possible. Condition **NC** would presumably say that utility functions are not interpersonally comparable but without specifying whether they are measurable on an ordinal, cardinal or some other scale. But how does one state this as an invariance requirement on the social welfare functional? Stegenga does not say. He would have done better to stick with Sen's well-understood framework rather that attempting to reformulate it.

The underlying problem is that Stegenga seems not to understand what interpersonal comparability of utility means in Sen's framework. In his Alexa and Beth example, he explains interpersonal comparability by saying that 'Alexa can transform her utility function only exactly as Beth does, and vice versa: if Beth multiplies each of her utilities by 2, then Alexa must do the same' (Stegenga 2014, p. ?). He then describes this as a 'needless constraint' on Alexa and Beth (ibid. p. ?). This is a confusion. In Sen's framework, interpersonal comparability of utility is not a constraint on the individuals, or on their utility functions, but on the aggregation rule. It does not say that Alexa 'can only transform' her utility function as Beth does, whatever that means. Rather it says (for the case of cardinal utility), that starting from a given profile of utility functions, if Alexa's and Beth's utility functions are subjected to the same positive linear transformation, then the resulting profile must be mapped to the same ordering of alternatives as the original profile.

In section 4 of his paper, Stegenga discusses my claim that in some cases at least, the simplicity of a theory or hypothesis may be measurable on a stronger than ordinal scale. He agrees that this is so, but observes that in other cases simplicity is only ordinally measurable. I agree with this; indeed I said as much in my own paper. Stegenga adds that if theories from very different areas of science were being compared, e.g. Bohr's theory of the atom with a particular statistical model, then ordinal comparison of their respective simplicity is likely all that could be achieved.[1] This may be correct but it is of little relevance, as such theories are not alternatives in the first place, so the issue of choosing between them does not arise.

Stegenga refers to simplicity and accuracy as 'theoretical virtues' rather than 'criteria of theory choice' as in my original discussion, which is harmless.

[1] Stegenga adds that it may 'turn out that the ordinal comparison is a weak ordering or an equivalence' (2014 p.?), which is a meaningless assertion. What he means to say is that the two theories may be ranked as equally simple.

However Stegenga introduces a terminological infelicity when he talks about the '*support* that a given theoretical virtue provides to a theory', and when he goes on to ask whether this 'support' is measurable on an ordinal or stronger scale. My original discussion made no mention of 'support' and simply asked whether a particular theoretical virtue itself, such as simplicity, is measurable on an ordinal or stronger scale. This is the clearest way to pose the issue.

Stegenga rightly observes that in a given context one theoretical virtue, such as simplicity, may be only ordinally measurable, while another, such as fit-with-the-data, may be measurable on a cardinal scale, for example. However he is too quick when he argues that in order to perform a comparison between two theoretical virtues (the analogue of the inter-personal comparisons of utility that are necessary to avoid Arrovian impossibility), they must be '*commensurable*—literally, they must share the same scale' (Stegenga 2014 p. ?). This is incorrect. If simplicity and fit-with-the data, for example, were both measured on their own ratio scales, so could each be rescaled independently of each other, this nonetheless brings with it a certain amount of comparability (as pointed out in Okasha 2011, p.101). Assertions such as 'in moving from theory T_1 to T_2, the percentage gain in simplicity is greater than the percentage loss in fit' become meaningful; and so long as both measurement scales only admit non-negative values, this amount of comparability is sufficient to avoid Arrovian impossibility, as shown by Tsui and Weymark 1997.[2] It is *not* necessary, in this example, that the two theoretical virtues be measured on the same scale.

If different theoretical virtues are measurable on different scale types, e.g. ordinal and cardinal, this is analogous to some individuals having ordinal utility functions and others cardinal, which gives an interesting twist to the aggregation problem. Stegenga says he 'knows of no work' in social choice theory that studies this issue (ibid. p.??). I have encountered two papers dealing with the issue (Khmelnitskaya 1996, Khmelnitskaya and Weymark 2000), containing results that could in principle be transposed to the context of scientific theory choice. More work on this issue might be interesting.

[2] Stegenga mentions this result on p.? of his article, but wrongly attributes it to Sen. He does not appreciate that the result disproves his assertion about 'commensurability'.

Stegenga's primary contention is that the comparability between theoretical virtues needed to avoid Arrovian impossibility will only rarely obtain, and thus that the 'enriched informational basis' escape route will only rarely be available. I see little real disagreement between us here. In my paper I argued that in some cases the Sen-style escape route does apply to theory choice while in others it does not. I gave two examples of theory choice rules where it does apply—Bayesianism and statistical model selection. Stegenga observes that these cases do not cover the whole of science, and that in other cases the escape route is not available. With that I agree; indeed it is essentially what I argued myself.

In his discussion of the non-comparability of the different theoretical virtues, Stegenga conflates two issues. In Okasha 2011, I observed that in statistical inference, a common measure of how well a theory fits the data is the 'sum of squares' (SOS) score, in which case the criterion 'fit-with-the-data' becomes measurable on the same scale-type as the dependent variable on the regression plot. Stegenga says that the SOS measure is 'entirely conventional' (ibid. p.??). This is something of an exaggeration (the SOS measure has many desirable properties), but it is true that other measures of fit exist, which need not be ordinally equivalent to the SOS score. However this point—that 'fit-with-the-data' may be measured in different ways—has nothing to do with the main point in Stegenga's paragraph, which is that, however it is measured, the permissible transformations of the fit-with-the-data scale are independent of the permissible transformations of the scales on which the other theoretical virtues are measured. This last point is correct, and is something I myself stressed; but it does not necessarily block the Sen-style escape route, for reasons given in my original paper and partly recapitulated three paragraphs back.

In Okasha 2011, I showed how the orthodox Bayesian approach to theory choice may be subsumed within a social choice-theoretic framework. There are two criteria of theory choice (or 'theoretical virtues'), prior probability $P(T_i)$ and likelihood $P(E/T_i)$, both of which are represented by real-valued functions on the theories $\{T_1,...,T_n\}$ that we wish to choose between, for a given body of evidence E. The Bayesian theory choice functional (the 'BCF' of my 2011 paper) then generates an overall ranking of the theories according to the value of the product $[P(T_i) \text{ x } P(E/T_i)]$. This theory choice rule satisfies the analogues of Arrow's non-domain axioms in the Sen-style framework, as I showed. Since probabilities are measured on an absolute scale (or if one regards the assignment of probability one to the certain event as conventional, on a ratio scale), the input to the theory choice rule contains of much more than merely ordinal non-comparable rankings. This neatly illustrates how Sen's escape route—enriching the informational basis—applies to theory choice.

Stegenga objects to this argument on the grounds that 'the Kuhnian theoretical virtues that Okasha began with...do not appear in his discussion of Bayesianism' (ibid p.??). He observes that Kuhn's criteria, such as simplicity and fruitfulness, are plausibly regarded as relevant to the determination of the prior probabilities that the Bayesian starts with. This is no doubt correct, but Stegenga's criticism is misplaced. As I made clear in my 2011 paper, my real interest was not so much in assessing Kuhn's own account of theory choice, which is rather idiosyncratic, but rather in the more general idea that theory choice in science is based on multiple criteria that may pull in different directions. As I observed, this latter idea is common to diverse philosophical approaches to scientific inference, including Bayesianism, inference to the best explanation, and statistical model selection, so is not specific to Kuhn. My point in discussing the Bayesian approach was not to endorse it, but rather to give a concrete illustration of how the informational enrichment strategy permits an escape from Arrovian impossibility. To criticise my discussion of Bayesianism on the grounds that it is not framed in terms of Kuhn's own criteria thus misses the dialectical point.

3. Morreau

Morreau (2014) applauds the general idea that scientific theory choice may profitably be compared with social choice and formalized similarly. However he argues that the two cases are not exactly analogous, and that no Arrow-style impossibility result threatens theory choice since Arrow's condition **U** (unrestricted domain) is inapplicable. Thus even if the input to theory choice rule consists solely of ordinal rankings, i.e. without informational enrichment, there is no threat of impossibility, Morreau argues.

In social choice, condition **U** says that the domain of the social choice rule is the set of all possible profiles of preference orders over the alternatives—the universal domain. This means that there are no *a priori* restrictions on the preferences that individuals are allowed to have: whatever their preferences, the social choice rule is required to output an overall ranking of the alternatives.

As Morreau notes, one of Arrow's original motivations for **U** was epistemic: we may want to design an aggregation procedure before we know what the actual preference profile is. Thus in an election, the rule for combining the voters' rankings of the candidates into an overall ranking should ideally be specified before the ballot opens, and thus before the actual preference profile is known. A different motivation for considering multiple preference profiles was given by Kolm (1996, 1997). He argues that even if the actual preference profile is known, to justify making a given social choice on the basis of this profile requires considering what social choice would have been made had the actual preference profile been different. So it is essential that the domain of the social choice rule contain multiple profiles.[3]

Whatever its justification, condition **U** evidently presupposes that the individuals in society could have had preferences different from the ones they actually do have. This seems unproblematic, at least in most cases. Suppose that three candidates are contesting a U.K. election, one from each of the main parties. One of the electors, an elderly man called Bob, has the following preference order: Labour $\succ$ Lib. Dem. $\succ$ Tory. It seems entirely conceivable that Bob could have had a different preference order. Bob might have undergone a rightwards shift in middle age and had exactly the opposite preferences. Or he might have long since tired of party politics and been indifferent among the three candidates. The same applies to the other electors too; so any profile of preference orders is possible. It is thus perfectly coherent, conceptually, to impose condition **U** on the aggregation rule in paradigmatic social choice problems.

Morreau argues that matters stand different with theory choice. He argues that some criteria of theory choice, such as simplicity, are 'rigid'. Suppose that theory T_1 is simpler than T_2 (by whatever yardstick of simplicity we are using). This fact could not have been otherwise, Morreau argues: there is no way that theory T_2 could have been simpler than T_1. Morreau's reason (in effect) is that the simplicity of a theory is an essential rather than an accidental property of it; and thus the relative simplicity of two theories is an essential property of the pair. For a

[3] See Weymark 2011 for useful discussion of Kolm's ideas.

given set of theories, there is only one possible way that they could have been ordered in terms of simplicity, namely the actual way. Therefore condition **U** is inapplicable, Morreau argues. It is unreasonable to demand that a theory choice rule be able to handle profiles that cannot possibly arise.

Morreau does not hold that all criteria of theory choice are rigid. He explicitly allows that 'accuracy' (or 'fit-with-the-data') is not rigid. Theory T_1 may fit the actual empirical data better than T_2, but had the data been different the reverse might have been true. So for a given set of theories, any possible ordering of those theories in terms of fit-with-the-data is conceivable. Thus it makes good sense to allow the domain of a theory choice rule to include profiles that order the theories by 'fit-with-the-data' in all possible ways. For as Morreau observes, we would ideally want our theory choice rule to be specified before the empirical data come in, and thus before we know what the 'fit-with-the-data' ordering is.

Morreau thus allows that the domain of a theory choice rule should contain multiple profiles; but he insists that the analogue of Arrow's condition **U** is too strong, since the rigidity of criteria such as 'simplicity' means that some profiles represent metaphysical impossibilities. Since Arrow derived his impossibility result with the help of condition **U**, Morreau concludes that no analogous result applies to theory choice: Arrow's theorem 'tells us precisely nothing' about the possibility of combining different criteria into a theory choice rule, he says (Morreau 2014, p.??).

Morreau's argument raises three issues. Firstly, are his claims about rigidity correct? Secondly, if so, does it follow that condition **U** is inapplicable to theory choice? Thirdly, if condition **U** is inapplicable, what are the implications for the rationality of theory choice? I address these issues in turn.

Morreau's claim that 'simplicity' is rigid seems quite right, at first blush. He argues that if Copernicus's theory was computationally simpler than Ptolemy's, then this could not have been otherwise; and if a particular statistical model contains fewer free parameters than another (so is simpler in that sense), this too could not have been otherwise. However there are potential counters to these claims.

Morreau's claims about rigidity are based on modal intuitions about how the essence / accident distinction applies to theories. Like any object, a scientific theory can presumably undergo some changes while retaining its identity. If

Newton had formulated his theory of gravity slightly differently, e.g. using a different calculus notation, it would still have been the same theory, intuitively. It may be difficult to say exactly which changes Newton's theory can undergo without ceasing to be the same theory. Had he proposed an inverse cube law this would presumably have been a different theory, not a variant of the same theory, but where do we draw the line? As for concrete objects, the essence / accident distinction is not easy to apply to scientific theories.

An interesting discussion of this point was offered by David Hull (1988), who argued that historical descent, rather than intrinsic properties, provides the key to individuating a scientific theory. Thus 'Darwinian theory', to use Hull's favourite example, denotes a lineage of ideas beginning with Darwin; these ideas may not share an intrinsic essence, but are united by descent. If this is correct, then the changes that a scientific theory, e.g. Copernicus's, can undergo while retaining its identity are presumably considerable. Whether these changes are sufficiently great so that the relative simplicity of Copernicus's and Ptolemy's theories could have been inverted I do not know; but the issue seems to me less clear-cut than Morreau assumes. (Note also that some philosophers regard appeals to simplicity in science as stemming from background empirical assumptions, rather than referring to intrinsic features of the theories themselves; this is argued persuasively by Sober 1988).

Despite the difficulties with individuating scientific theories, I am inclined to agree with Morreau's claim that some criteria of theory choice, such as simplicity, are rigid, as I largely share his modal intuitions. So let us grant this point. Does it follow that Arrow's condition **U** (unrestricted domain) is inapplicable to theory choice? Morreau seems on strong ground here. If the relative simplicity of Copernicus's and Ptolemy's theories could not have been other that it is, surely we should not require of a theory choice rule that it accept as input profiles in which the simplicity ordering of these two theories is different from the actual one, as *ex hypothesi* they cannot arise?

In arguing this point Morreau stresses that in standard Arrovian social choice theory there is a fixed set of individuals and a fixed set of alternatives; the different profiles in the domain are different possible orderings by the individuals of the alternatives. The individuals and the social alternatives do not vary across profiles. The same should apply to theory choice, if it is to be modelled on social choice. Thus if the set of alternatives includes theories T_1 and T_2, and if T_1 is in fact simpler than T_2, then any profile in the domain in which T_2 occurs higher up the simplicity ordering than T_1 describes a hypothetical scenario in which T_2 itself is simpler than T_1 itself; *not* a scenario in which some *other* pair of theories are so related. Morreau is right to stress this point, which I admit I overlooked when I argued that condition **U** as applied to theory choice is unexceptionable. This is a well-taken criticism.

One possible response might be to construe the 'alternatives' as abstract labels, denoting items whose identity is not necessarily fixed as we move from profile to profile. In social choice terms, this would be to suggest that the different profiles in the domain refer to patterns of preferences that the individuals might have had over different items bearing the same 'labels'. This suggestion has occasionally been mooted in the social choice literature, precisely to avoid the type of objection that Morreau is making, for example by Blackorby, Donaldson and Bossert (2006 p. 281). However in a recent paper, which complements his reply to my article, Morreau (2013) is sharply critical of treating the alternatives as 'labels', arguing that this constitutes a serious modification of the standard social choice framework and is a recipe for confusion. I agree with this assessment.

Let us grant then that some criteria for theory choice are rigid, and that Arrow's condition **U** as applied to theory choice is inappropriate. What follows? Morreau concludes that Arrow's theorem does not apply to theory choice, so there is no threat to the rationality of science. Because of rigidity, and the consequent inapplicability of condition **U**, Arrow's theorem 'gets no grip' he argues (ibid. p. ??), and so the 'impossibility scare' can be seen off (ibid. p. ??) . However this is too quick.

As is well-known, Arrow's condition **U** is actually much stronger than is needed to derive his impossibility result. Since the late 1970s, an extensive research program has investigated whether an impossibility result can be derived with weaker domain assumptions; the answer turns out to be yes.[4] Central to this literature is the notion of an 'Arrow-inconsistent domain', which refers to any subset of the universal domain on which Arrow's axioms **N**, **P**, and **I** are jointly unsatisfiable. (Recall that the universal domain is the set of all profiles of

[4] See Gaertner 2001 or Le Breton and Weymark 2011 for overviews of this work.

preference orders over the alternatives.) We know from Arrow's own work that the universal domain itself is Arrow-inconsistent, but so too are many proper subsets of the universal domain. Thus one cannot argue that because condition **U** is inapplicable in a certain circumstance, that the threat of impossibility is thereby eliminated.

How does this bear on the theory choice case? To sharpen the issue, suppose we have a finite set of alternative theories $\{T_1,\ldots,T_n\}$. Let R be the set of all orderings of these n theories. To keep things tractable, suppose we are using just two criteria of theory choice: simplicity and accuracy; this involves no serious loss of generality. Let us assume with Morreau that simplicity is rigid. Let $R_s \in R$ denote the sole admissible simplicity ordering. We agree that accuracy is completely non-rigid; so all elements of R represent admissible accuracy orderings. The domain of the theory choice rule is therefore the Cartesian product R_s x R, which is a proper subset of the universal domain R^2. We know that R^2 is Arrow-inconsistent, but what about R_s x R? That is the crucial question.

I do not know the answer to this question; so far as I am aware this case is not covered by any of the extant theorems on domain-restriction in social choice.[5] Morreau's paper does not answer the question either. So when he says, in relation to my statistical model selection example (which precisely involves two criteria, accuracy and simplicity, one of which is rigid and the other not), that 'certainly Arrow's theorem doesn't limit the possibilities for choosing among models in this example' (ibid. p.??), this is rather misleading. In the absence of a proof that the domain R_s x R is Arrow-consistent, the right conclusion to draw is that we do not know whether an Arrovian impossibility result applies in this case or not. There is an unresolved mathematical question here.[6] So while Morreau's assertion may be

[5] Most theorems in this literature identify only sufficient conditions for a domain to be Arrow-inconsistent, and / or operate with so-called 'economic domains' which contain more structure than the abstract sets of alternatives that Arrow worked with, and / or require that the admissible preference orderings for each individual be the same. See Gaertner 2001 or Le Breton and Weymark 2011.

[6] John Weymark (personal communication), building on a suggestion of Wulf Gaertner, has partially resolved this question. He has shown by example that in the special case where both criteria *strictly* order the alternative theories, the domain R_s

literally true, in that Arrow's original theorem (which used condition **U**) does not apply, the real question is whether an Arrow-style theorem with a weaker domain assumption applies, and this has not been settled.

This brings me to my only real complaint with Morreau's paper, which is that in a number of places he implies that theory choice faces no threat of impossibility on the grounds that condition **U** does not apply. (For example, he says that the analogy with social choice 'looks harmless' for the rationality of science (p.??) and that his analysis has put 'the impossibility scare safely behind us' (p.??).) This may be true, but Morreau has not shown it. To do so, he would need to show that with the appropriate domain assumption, e.g. R_s x R in the example above, there exist theory choice rules that satisfy conditions **N**, **P** and **I**. One way to show this would be to provide an example of a theory choice rule that satisfies these conditions on the domain in question. Morreau does not do this, so his confidence that the threat of impossibility has been allayed is misplaced. It remains open whether an impossibility result holds for theory choice or not, even granting Morreau his points about rigidity and domain restriction.

To be fair to Morreau, he is well aware that condition **U** is stronger than needed to derive Arrow's result, and indeed says so explicitly on page ? of his paper. Moreover he adds, in a footnote, that it is an 'open question' whether with some rigid criteria, the domain of the theory choice rule will still contain diverse enough profiles to allow an Arrow-style impossibility result to go through (p. ?, fn 12). This is exactly right. But given that the question is open, Morreau is not entitled to suggest that his analysis 'vindicates Kuhn' as per his title, nor that he has put the impossibility scare behind us. A successful vindication of Kuhn, that is, of Kuhn's claim that there are many acceptable algorithms for theory choice,

---

x R is Arrow-consistent. However Weymark also shows that in this special case, for any triple of alternatives there is at least one pair on which one of the criteria is a dictator, i.e. the overall choice for that pair is determined by a single criterion. Though not as bad as Arrovian dictatorship, this is nonetheless an undesirable feature. For details of Weymark's example and proof, please contact the author.

would at the very least have to show that the typical domain for theory choice is Arrow-consistent, which Morreau does not do.

This complaint is relevant to Morreau's critique of the 'informational enrichment' escape route as applied to my statistical model selection example. In that example, I identified a domain restriction that arises because one criterion (accuracy) can only take negative values, while the other (simplicity) can only take positive integer values. I claimed that this domain restriction alone does not suffice to alleviate Arrovian impossibility, and sketched a proof.[7] Morreau agrees that this is so, writing 'Okasha's mild domain restriction is not crucial....it by itself will not alleviate the Arrow impossibility' (ibid p. ?). He continues 'but there is another domain restriction in force as well', that arises because simplicity is rigid. Morreau here creates the impression that this additional domain restriction *is* crucial, and *does* block the impossibility. This may be so; but it is a conjecture, not something that he has shown.

Note also that Morreau's critique of my use of Arrow's framework to model scientific theory choice does not apply to my Bayesian example. (This example is briefly described above in my response to Stegenga). In this case the two criteria—prior probability and likelihood—are not rigid. Clearly, a given set of theories could be ordered by prior probability in any way, at least on a subjective interpretation of probability[8]; and the same is true of the likelihood ordering, given that the empirical data might have been different. So in this case, the theory choice rule (the Bayesian theory choice functional) is able to avoid impossibility not because of domain restriction but because of informational enrichment, as proved in the Appendix to my original article.

To conclude, Morreau, Stegenga and I agree that scientists do choose between rival theories based on how those theories score against multiple criteria, and seem able to do this in a rational way. Given Arrow's theorem, a question

[7] More accurately, I provided a full proof of the analogous claim in relation to my Bayesian example, and argued that a similar proof could be given for the statistical model selection case. See Okasha 2011 p. 100, fn. 22, and Appendix.

[8] Here I assume that none of the theories logically implies any of the others. This is reasonable as otherwise the theories would not constitute genuine alternatives in the first place.

arises as to how this could be. There are a number of possible explanations. One is that theory choice in science is often binary, i.e. the choice is between a pair of theories. Another is informational enrichment: the input to the theory choice rule may consist of more than merely ordinal non-comparable rankings. Another is domain restriction: the appropriate domain of the theory choice rule may be sufficiently small to be Arrow-consistent (as Morreau hopes). It would be a mistake, in my view, to insist that any one of these explanations provides the full resolution of the puzzle. There may be an element of truth in each of them.[9]

[9] Thanks to Michael Morreau, Kit Patrick, Wulf Gaertner, and John Weymark for discussion and email correspondence. This work was supported by the European Research Council Seventh Framework Program (FP7/2007–2013), ERC Grant agreement no. 295449.

References

Arrow, K. J., A. K. Sen and K. Suzumura (eds) 1996: *Social Choice Re-Examined, Volume 2*. London: Macmillan Press.

Arrow, K. J., A. K. Sen and K. Suzumura (eds) 2011: *Handbook of Social Choice and Welfare, Volume 2*. North-Holland: Elsevier.

Arrow, Kenneth 1951: *Social Choice and Individual Values*. New York: John Wiley.

Blackborby, C., D. Donaldson and W. Bossert 2006: 'Anonymous Single-Profile Welfarism'. *Social Choice and Welfare,* 2006, 27, pp. 279–87.

Fleurbaey, M., M. Salles and J. A. Weymark (eds) 2011: *Social Ethics and Normative Economics*. Berlin: Springer-Verlag.

Gaertner, Wulf 2001: *Domain Conditions in Social Choice Theory*. Cambridge: Cambridge University Press.

Hull, David L. 1988: *Science as a Process*. Chicago: University of Chicago Press.

Khmelnitskaya, Anna B. 1996: 'Social Choice Problems with Different Scales of Individual Welfare Measurement for Different Subgroups of Individuals'. In Kleinschmidt, Bachem, Derigs, Fischer, Leopold-Wildburger and Möring 1996, pp. 252–7.

Khmelnitskaya, Anna B. and John A. Weymark 2000: 'Social Choice with Independent Subgroup Utility Scales'. *Social Choice and Welfare,* 17, pp. 739–48.

Kleinschmidt P., A. Bachem, U. Derigs, D. Fischer, U. Leopold-Wildburger and R. Möring (eds) 1996: *Operations Research Proceedings 1995*. Berlin: Springer-Verlag.

Kolm, Serge-Christophe 1996: 'Rational Just Social Choice'. In Arrow, Sen and Suzumura 1996, pp. 167–96.

Kolm, Serge-Christophe 1997: *Modern Theories of Justice*. Cambridge MA: Cambridge University Press.

Kuhn, Thomas 1969: *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, Thomas 1977a: ‘Objectivity, Value Judgment, and Theory Choice’. In Kuhn 1977b, pp. 320–39.

Kuhn, Thomas 1977b: *The Essential Tension*. Chicago: University of Chicago Press.

Le Breton, Michel and John A. Weymark 2011: ‘Arrovian Social Choice Theory on Economic Domains’. In Arrow, Sen and Suzumura 2011, pp. 191–299.

Morreau, Michael 2013: ‘Mr. Fit, Mr. Simplicity and Mr. Scope: From Social Choice to Theory Choice’, *Erkenntnis* 2013. ISSN 0165-0106.s doi: 10.1007/s10670-013-9549-x.

Morreau, Michael 2014: ‘Theory Choice and Social Choice: Kuhn Vindicated’, *Mind*

Okasha, Samir 2011: ‘Theory Choice and Social Choice: Kuhn versus Arrow’, *Mind*, 477, pp. 83–115.

Sen, Amartya 1970: *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Sen, Amartya 1977: ‘On Weights and Measures: Informational Constraints in Social Welfare Analysis’. *Econometrica*, 45, pp. 1539–72.

Sober, Elliott 1988: *Reconstructing the Past.: Parsimony, Evolution and Inference*. Cambridge MA: MIT Press.

Stegenga, Jacob 2014: ‘Theory Choice and Social Choice: Okasha versus Sen’, *Mind*

Tsui, Kai-yuen and John A. Weymark 1997: ‘Social Welfare Orderings for Ratio Scale Measurable Utilities’. *Economic Theory*, 10, pp. 241–56.

Weymark, John A. 2011: ‘On Kolm’s Use of Epistemic Counterfactuals in Social Choice Theory’. In Fleurbaey, Salles and Weymark 2011, pp. 279–301.