Theory Choice and Social Choice: Kuhn versus Arrow

SAMIR OKASHA

Abstract

Kuhn's famous thesis that there is 'no unique algorithm' for choosing between rival scientific theories is analysed using the machinery of social choice theory. It is shown that the problem of theory choice as posed by Kuhn is formally identical to a standard social choice problem. This suggests that analogues of well-known results from the social choice literature, such as Arrow's impossibility theorem, may apply to theory choice. If an analogue of Arrow's theorem does hold for theory choice this would refute Kuhn's thesis, but it would also pose a threat to the rationality of science, a threat that is if anything more worrying than that posed by Kuhn. Various possible 'escape routes' from Arrow's impossibility result are examined, in particular Amartya Sen's idea of 'enriching the informational basis'. It is shown that Sen's idea can be applied to the problem of theory choice in science. This in turn sheds light on two well-known approaches to inductive inference in philosophy of science: Bayesianism and statistical model selection.

1. Introduction

In the Postscript to *The Structure of Scientific Revolutions*, Thomas Kuhn famously argued that there is 'no neutral algorithm for theory choice' in science. Kuhn allowed that scientists might have good reasons for choosing one theory over its rivals, citing 'accuracy, simplicity, fruitfulness and so on' as examples of such reasons, but he insisted that they fall short of providing an algorithm (Kuhn 1969 p. 199). Even if two scientists agree on the features that a good theory should have, they will not necessarily be led to make the same choices, Kuhn argued, for they may weight the features differently. For example, the two scientists might agree that accuracy and simplicity are both important theoretical virtues, but disagree about their relative importance, and thus be led to choose different theories. Neither can be called irrational, Kuhn insisted, and neither was necessarily acting unscientifically.

The idea that there is no algorithm for theory choice met with a favourable response from Kuhn's critics, even among those unsympathetic to other aspects of his

philosophy, and seems to be regarded as fairly uncontroversial.[1] In part, this is probably because Kuhn's 'no algorithm' claim tallies well with the widely-held view that Carnap-style inductive logic is an impossible dream. Carnap's aim was precisely to devise an algorithm for inductive reasoning—a rule that, given a set of hypotheses and a body of data as input, would tell us which hypothesis is best confirmed by the data. The perceived failure of Carnap's project, which Kuhn himself alluded to in a later essay[2], lent credence to his claim that there is no algorithm for theory choice.

My aim in this paper is to explore Kuhn's 'no algorithm' thesis from a new angle, by drawing on some ideas from social choice theory, in particular Arrow's famous impossibility theorem (Arrow 1951). Social choice theory studies the problem of how a society should choose between a set of 'social alternatives', given that the individuals in society may have different preferences over those alternatives. I show that the problem of theory choice, as described by Kuhn, is formally identically to a standard social choice problem. This raises an interesting possibility: using formal results from social choice theory, such as Arrow's theorem, to assess Kuhn's claim that there is no algorithm for theory choice. That is my primary goal in this paper; a secondary goal is to illustrate more generally how techniques from theoretical economics can be applied to problems in epistemology.

The structure of this paper is as follows. Section 2 examines and clarifies Kuhn's 'no algorithm' argument and its philosophical consequences. Section 3 introduces social choice theory and gives a non-technical exposition of Arrow's theorem. Section 4 shows how the problem of theory choice can be cast in a social choice framework, and asks whether an analogue of Arrow's theorem applies to theory choice. Section 5 examines possible 'escape routes' from Arrow's theorem. Section 6 discusses the most promising escape route, namely Amartya Sen's 'informational basis' approach. Section 7 applies Sen's approach to the problem of theory choice. Section 8 shows how the foregoing results shed light on two well-known approaches to inductive inference in philosophy of science: Bayesianism, and statistical model selection. Section 9 draws together the pieces and concludes.

[1] See for example the discussion in W. H. Newton-Smith 1981.

[2] In that essay Kuhn wrote 'most philosophers of science would … I think, now regard the sort of algorithm which has traditionally been sought as a not quite attainable ideal' (Kuhn 1977a, p. 328).

2. Kuhn's 'no algorithm' argument: clarifications

Kuhn developed his 'no algorithm' argument most thoroughly in a 1977 essay entitled 'Objectivity, Value Judgment and Theory Choice'. In that essay, he identifies five criteria that provide 'the shared basis for theory choice', namely accuracy, consistency, scope, simplicity, and fruitfulness (Kuhn 1977a, p. 321). These five, he says, are 'the standard criteria for evaluating the adequacy of a theory', widely agreed on by mainstream philosophers of science. Kuhn has no quarrel with the standard view that these criteria play a key role in scientific theory choice; indeed, he regards them as partially constitutive of what science is. However he argues, using examples from the history of science, that the criteria fail to uniquely determine theory choice, for two reasons. Firstly, the criteria are ambiguous—it may be unclear which of two theories is simpler, for example. In some respects Copernicus' theory was simpler than Ptolemy's, Kuhn says, but in others it was not. Secondly, there is the problem of how to appropriately weight the criteria, when they pull in different directions. How should simplicity be traded off against accuracy and scope, for example? Kuhn says that 'no progress' has been made towards solving this problem (Kuhn 1977a, p. 329).

It is this second problem—weighting—that will be the focus of our attention here, so the first problem—ambiguity—will be ignored. In any case, the first problem arguably collapses into the second. Consider Kuhn's own example. He says that Copernicus's theory was simpler than Ptolemy's in that it invoked more parsimonious mathematics, but was no simpler in that the computational labour required to predict planetary positions was the same for both. If this is correct, then simplicity, in this example, needs to be sub-divided into two criteria: mathematical parsimony and computational ease, neither of which is ambiguous. Of course, this raises the question of how the two types of simplicity should be weighted, which is more important, etc. But that is just an instance of the second problem. It seems, therefore, that the ambiguity of Kuhn's five criteria provides no principled reason to doubt the existence of an algorithm for theory choice. Disambiguation can always be carried out by sub-dividing an ambiguous criterion, though this exacerbates the weighting problem. So the latter is more fundamental.

Suppose it is true that there is no algorithm for theory choice. What follows? The conclusion Kuhn drew is that *value judgements* play an inevitable role in theory choice, and thus that the ideal of 'objectivity', as that notion was understood by traditional philosophers of science, is unattainable. This does *not* mean that theory

choice is irrational, Kuhn stressed, or that 'anything goes', but rather that the traditional conception of rationality is too demanding. Two scientists, on the basis of the same empirical evidence, could arrive at different theories without either of them being irrational, Kuhn argues. So the 'no algorithm' argument does not undermine the rationality of science, he thinks, but rather forces us to a more realistic conception of what rational theory choice is like. Whether or not we accept this, Kuhn is surely right that the 'no algorithm' argument has important epistemological consequences.

Importantly, when Kuhn says there is no algorithm for theory choice, he means that there is no *unique* algorithm. His point is that given the five criteria—simplicity, accuracy, fruitfulness etc.—which all parties agree form the shared basis for theory choice, there are many conceivable algorithms that one could construct, and no obvious way of choosing between them. (Thus Kuhn talks about 'the algorithms of different individuals' in a scientific community, while insisting that there is no such thing as 'the algorithm of objective choice' (Kuhn 1977a, p. 328).) So the problem is that there are *too many* algorithms for theory choice, each perfectly acceptable, and no way of singling out the 'right' one. This is why Kuhn sometimes expresses his thesis by saying there is no *neutral* algorithm for theory choice (Kuhn 1969, p. 199).

My point in stressing this is that there is another, quite different thing that might be meant by saying there is 'no algorithm' for theory choice. One might mean that there is no acceptable algorithm at all, not that there are too many. It might be the case that there is no way of constructing an algorithm, based on the five criteria, which meets minimal standards of acceptability. But this is not Kuhn's claim. Rather, he thinks that there are *many* acceptable algorithms, each of which weights the five criteria differently, and each as rationally defensible as each other.

Another way to capture the distinction is this. Kuhn will presumably agree that there are certain minimal standards which any acceptable algorithm for theory choice must meet. For example, if theory T_1 scores higher than T_2 on *each* of his five criteria, then an algorithm that selects T_2 over T_1 is obviously unacceptable—no rational scientist could use it. So these minimal standards, whatever exactly they are, enable us to rule out some possible algorithms for theory choice. But on Kuhn's view, many algorithms will still remain—and there is no way of narrowing down the choice to a single one. An alternative view, however, is that *no algorithms at all* will remain, after those that fail to meet the minimal standards have been discarded. These are

quite different views, and are mutually incompatible, though each might be expressed by saying that there is 'no algorithm' for theory choice.

The distinction between these two views might be thought inconsequential, for both imply that there is no *unique* algorithm for theory choice, though for different reasons. And if Kuhn is right that the notion of rationality of traditional philosophy of science presumes the existence of a unique algorithm, then there is a threat to that notion in either case. This may be so; but it is still surely important to know *why* there is no unique algorithm for theory choice, if there isn't. Is it because there are many acceptable algorithms and no good way of choosing between them, or because there are no acceptable algorithms? The epistemological consequences might be similar in either case, but the distinction is a real one and will play an important role in what follows.

The idea that theory choice is based on multiple criteria, which may pull in different directions, is not unique to Kuhn's philosophy of science. Rather, it is common to diverse philosophical views on how scientific inference works. For example, proponents of 'inference to the best explanation' cite multiple factors, such as simplicity, unifying power, and scope, which enter into the assessment of how good a candidate explanation is (Thagard 1978, Lipton 1990). Bayesians argue that the choice between rival theories depends on how well they score on two different criteria—prior probability and likelihood—which can conflict (Howson and Urbach 1992, Earman 1992).[3] Finally, proponents of 'statistical model selection' argue that the choice between rival hypotheses again depends on two factors—fit-with-the-data and simplicity—which typically do pull in different directions (Forster and Sober 1994, Forster 2001). So the issues to be discussed below have a relevance that extends beyond an assessment of Kuhn's 'no algorithm' thesis.

3. Social choice theory and Arrow's impossibility theorem

Social choice theory deals with the problem of aggregating individuals' preferences, over a set of alternatives, into a single 'social preference'. For example, suppose a given society has four alternatives: building a school, a hospital, an airport, or a cinema, only one of which can be chosen. Each individual in society is assumed to have a *weak preference order* over the alternatives, that is, a ranking of the

[3] The point being that $P(H_1/e) > P(H_2/e)$ if and only if $P(H_1).P(e/H_1) > P(H_2).P(e/H_2)$. See Sect. 8 below.

alternatives from best to worst, with ties permitted. Formally, a weak preference order is a binary relation that is transitive, reflexive and complete. The preference order of the i^{th} individual will be denoted R_i, so 'xR_iy' means that the i^{th} individual weakly prefers alternative *x* to alternative *y*, that is, she doesn't strictly prefer *y* to *x*. From the weak preference relation R_i, we can define a corresponding relation of strict preference P_i, and of indifference I_i.[4]

Suppose that for each of the *n* individuals in society, we know their preference order over the four alternatives. We can encode all this information in a single *profile*, denoted $<R_1, \ldots, R_n>$, which is simply a list, or vector, of preference orders, one for each individual. An example of a profile is contained in Table 1 below. Given this information, we would then like to be able to construct a single social preference order R, which ranks the four alternatives in terms of how good they are 'for society as a whole'.[5] On any reasonable ethical view, the social preference order should depend somehow on the preference orders of the individuals in society, but how exactly?

John	1. Cinema	2. School	3. Hospital	4. Airport
Mary	1. Hospital	2. Cinema	3. Airport	4. School
Jane	1. School	2. Airport	3. Hospital	4. Cinema
Peter	1.Cinema	2. Airport	3. Hospital	4. School

Table 1: A Profile of Preference Orders for Four Individuals

In his seminal 1951 work, Kenneth Arrow devised a novel way to study this problem. Arrow's idea was to consider a function from profiles of individual preference orders to a social preference order, and then ask what conditions the function should satisfy. I will call such a function a 'social choice rule'.[6] The rule takes as input a profile of preference orders, and yields as output a single social preference order. In other words, given the preferences of all the members of society over the alternatives, the rule tells us what the social preference order should be. A

[4] Thus xP_iy iff xR_iy and it is not the case that yR_ix, while xI_iy iff xR_iy and yR_ix

[5] A social preference order is an object of the same sort as an individual preference order, i.e. a transitive, reflexive, complete binary relation over the alternatives.

[6] Arrow himself used the expression 'social welfare function', but this is often used in another sense today.

social choice rule is thus a kind of algorithm for making social choices, based on information about individuals' preferences.

Arrow proposed four conditions that any reasonable social choice rule should satisfy: unrestricted domain (**U**), weak Pareto (**P**), independence of irrelevant alternatives (**I**), and non-dictatorship (**N**). He then proved, remarkably, that there exists no social choice rule that satisfies all four conditions, so long as there are at least three social alternatives. That is the content of his famous 'impossibility theorem'. If we agree with Arrow that all four conditions are reasonable ones, his result spells bad news for the possibility of making coherent social decisions.

Condition **U** says that the domain of the social choice rule is the set of all possible profiles. This means that whatever the preferences of the individuals in society, the rule must output a social preference order, that is, there is no a priori restriction on the preferences that individuals are allowed to have. In many applications of Arrow's framework, this condition is extremely natural. For example, if the 'alternatives' are candidates in an election, then condition **U** says that voters can rank the candidates however they like, and the rule must still output a social preference, that is, an election result. This is obviously reasonable. In other applications, restricting the domain of the social choice rule may make sense; for example, if the alternatives are different ways of dividing up society's resources, then the assumption that individuals prefer more to less, ceteris paribus, suggests a natural domain restriction. But in general, condition **U** is well-motivated.

Condition **P**—weak Pareto—says that if all individuals in society strictly prefer alternative *x* to *y*, then society should also prefer *x* to *y*, that is, the social preference order must rank *x* above *y*. This seems indisputable: if everyone would rather have a cinema to a swimming-pool, then 'cinema' should obviously be higher than 'swimming pool' in the social ranking. This captures the intuitive idea that social choices must reflect what the members of society want; so if they all want the same thing, that is what should be chosen.

Condition **N**—non-dictatorship—says that there cannot be an individual who is such that whenever he or she strictly prefers alternative *x* to *y*, so does society. Such an individual would be a dictator—their preferences would automatically over-ride those of all other members of society. The existence of a dictator is clearly undesirable, as it conflicts with basic democratic ideals. So condition **N** seems unexceptionable.

Condition **I**—independence of irrelevant alternatives—is at the crux of Arrow's argument, and is slightly trickier than the others. It says that the social choice between alternatives *x* and *y* can only depend on individuals' preferences between *x* and *y*—not on their preferences over other alternatives. More precisely, consider two profiles of individual preference orders $<R_1, \ldots, R_n>$ and $<R'_1, \ldots, R'_n>$, such that for every individual i, xR_iy if and only if xR'_iy, that is, each individual's preference for *x* over *y* is the same in the two profiles. Condition **I** then says that the social choice rule, when applied to both profiles, must yield the same social preference for *x* over *y*, that is, *x*R*y* if and only if *x*R′*y*. Any differences between the two profiles are irrelevant to the social choice between *x* and *y*, according to condition **I**, since the two profiles are identical in the only respect that matters.

The intuitive force of condition **I** can be seen by considering an election in which voters must rank three candidates, Labour, Tory and Liberal, in order of preference. Various different ways of aggregating the individual preferences into a single social preference are conceivable. Condition **I** imposes a requirement on acceptable aggregation schemes—it says that the social preference between the Labour and Tory candidates, for example, can depend only on the individuals' preferences between Labour and Tory. This is highly intuitive— in order to determine whether the Labour or Tory candidate is socially preferable, surely the individuals' preferences involving the *Liberal* candidate should not matter? If you know of each individual whether they prefer the Labour to the Tory candidate (or are indifferent), then you know everything that is relevant to determining the social choice between these two candidates, according to condition **I**.

Arrow wrote that his four conditions 'taken together, express the doctrines of citizens' sovereignty and rationality in a very general form' (Arrow 1951, p. 31). Condition **I** is in fact somewhat controversial, as discussed below, but nonetheless, Arrow's four conditions arguably represent quite reasonable constraints on a social choice rule. But remarkably, Arrow proved that all four cannot be simultaneously satisfied, so long as there are at least three alternatives; equivalently, any social choice rule that satisfies conditions **U**, **I**, and **P** must be a dictatorship of one individual.

Arrow could have expressed his impossibility result by saying that there is 'no algorithm' for social choice that meets certain reasonable conditions. This way of

expressing Arrow's theorem immediately suggests a comparison with Kuhn's views on theory choice, to which I now turn.

4. Theory choice cast as a social choice problem

The problem of theory choice, as formulated by Kuhn, hinges on the fact that there are multiple desiderata that we want our theories to satisfy—simplicity, accuracy, scope etc. A theory that scores well on one desideratum might score badly on another, hence the weighting problem that Kuhn discusses. This problem may seem quite different to the social choice problem as formulated by Arrow, but in fact the two share a common structure. The key to seeing this is to regard each criterion of theory choice as an 'individual', with their own 'preference order' over the alternative theories. This may sound odd, but can be easily explained.

Take for example simplicity. Let us assume that simplicity can be defined reasonably precisely, enough to permit pair-wise comparisons between the theories that we wish to choose between.[7] Then, we can define a binary relation 'is at least as simple as', on the set of alternative theories, which will be a weak ordering, that is, reflexive, transitive, and complete. Let us do the same for accuracy, scope, and the other Kuhnian criteria. From a formal point of view, each criterion is then analogous to an individual in Arrow's set-up. Just as each individual rank-orders the social alternatives, according to how much they like them, so each criterion rank-orders the alternative theories, according to how well they satisfy it. So each of Kuhn's criteria corresponds to an individual in Arrow's framework, and the alternative theories correspond to the social alternatives.

It might be objected that for some criteria of theory choice, the binary relation will not be complete. Take for example scope. Plausibly, one might take a theory's 'scope' to be its total set of logical consequences, and the relation 'T_1 has at least as much scope as T_2' to mean that T_2's consequence class is a subset of T_1's. But this relation, though reflexive and transitive, need not be complete, for the consequence classes of a pair of theories may be non-nested, that is, the theories may be non-comparable for scope. Though this is a valid point, 'scope' is arguably the only one of Kuhn's five criteria that it affects. (In the case of simplicity, for example, it is plausible that for any two theories, either one is simpler than the other or they are

[7] This assumption is not unproblematic, but the problems it raises are orthogonal to those under discussion here.

equally simple, i.e. 'is at least as simple as' is complete.) So the completeness assumption can be justified as a reasonable idealization. After all the assumption that individuals' preference relations are complete is also an idealization.[8]

The next step is to consider a 'theory choice rule', defined by direct analogy with an Arrovian social choice rule. Given a profile of weak orders, one for each criterion of theory choice, a theory choice rule yields a single ordering of the alternative theories. So for example, suppose we have four alternative theories, and three criteria: simplicity, accuracy and scope. By assumption, we know how to rank-order the theories by each criterion. We feed this information into the theory choice rule, which then outputs an 'overall' ranking of the theories, from best to worst. Formally, the theory choice rule is defined in exactly the same way as Arrow's social choice rule.

Next, let us ask whether Arrow's four conditions apply to the theory choice rule. Condition **U** (unrestricted domain) 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 a priori restriction on the permissible rankings that are fed into the rule. Such a restriction *might* make sense if there is an intrinsic trade-off (or correlation) between two of the criteria. For example, if greater simplicity always involves a sacrifice of accuracy, then the simplicity rank-ordering will be the inverse of the accuracy rank-ordering. This will rule out some possible inputs to the rule, which implies a natural domain restriction. But unless we have specific reason to think such trade-offs must always obtain, condition **U** seems reasonable.

Condition **P** (weak Pareto) seems undeniable. If theory T_1 does better than theory T_2 by each of Kuhn's criteria, that is, it is simpler *and* more accurate *and* more fruitful etc., then it must surely be preferred overall. This seems as obvious as its analogue for social choice. What about condition **N** (non-dictatorship)? It says that there is no one criterion such that if T_1 is ranked above T_2 by that criterion, then T_1 is

[8] There is in fact a technical trick to get around the problem. Suppose R is reflexive and transitive but incomplete. We can then extend R to a complete relation R*, by stipulating that for any two objects *x* and *y* that are not related by R, *x*I**y*, i.e. neither *x*R**y* nor *y*R**x*. The relation R* will then be reflexive and complete, but non-transitive; however, it will be *quasi-transitive* (which means that P*, the corresponding strict preference relation, is transitive, but I* is not) (Sen 1969). Arrow's theorem will then apply; for the theorem does not in fact require that the individual preference orders be fully transitive – quasi-transitivity is enough. (By contrast, the full transitivity of the social preference order is essential to the theorem.)

automatically above T_2 in the overall ranking. This condition makes good sense, so long as we agree that all the criteria are relevant to theory choice. Violation of the condition would mean that one criterion, for example simplicity, was regarded as so important that a less simple theory would *never* be preferred to a more simple one, however highly it scored on the other criteria.

What about condition **I** (independence of irrelevant alternatives)? It says that the overall ranking of T_1 and T_2 should depend only on how the criteria rank T_1 and T_2, not on how they rank other theories. So for example, suppose we have three criteria, simplicity, accuracy and scope, and two theories. Suppose T_1 is simpler than T_2, T_2 is more accurate than T_1, and T_1 has greater scope than T_2. Condition **I** says that this is all the information that is relevant to the overall ranking of T_1 versus T_2; so if in this case the theory choice rule ranks T_1 above T_2 (for example), then it must rank T_1 above T_2 in every relevantly similar case, that is, every case where T_1 is simpler than T_2, T_2 more accurate than T_1, and T_1 greater in scope than T_2. As with the social choice rule, this condition has strong intuitive appeal, capturing the idea that rational theory choice shouldn't depend on irrelevant factors.

If we agree that **U**, **P**, **N** and **I** are conditions on reasonable theory choice, then it is obvious that an Arrovian impossibility result applies. So long as there are at least three alternative theories, there exists no theory choice rule that satisfies all four conditions. This spells bad news for the possibility of making 'rational' theory choices.

One might naturally express this impossibility result by saying that there can be 'no algorithm' for rational theory choice. This *sounds* similar to what Kuhn said, but recall the discussion of section 2. As we saw, Kuhn meant that there is no *unique* algorithm; he argued that the multiple criteria for theory choice could be combined into a decision rule in *many* ways, and there is no good way of choosing between them. So the problem according to Kuhn is that there are *too many* algorithms. But the Arrow impossibility result implies the opposite—there is *no* algorithm for theory choice that meets reasonable conditions. When Kuhn saw an embarrassment of riches, Arrow tells us that there is nothing at all.

Both Kuhn's view and the Arrow-inspired view imply, obviously, that there is no *single* algorithm for theory choice, over three or more alternatives, which is rationally acceptable. This conclusion conflicts with the traditional ideal of rationality, associated with Carnap, according to which two rational agents with the same 'total

evidence' must end up in the same epistemic state. In the context of theory choice, the Carnapian ideal implies that if two scientists agree on what the relevant criteria are, and agree about how well each theory performs on each criterion, then they should agree on how to rank the theories. On Kuhn's view, such agreement is unattainable—the two scientists may weigh the criteria using different algorithms, and there is no saying which is rationally correct. On the Arrovian view, agreement is also unattainable, but for a very different reason, namely that *no* algorithm meets minimal standards of rationality.

Despite both implying that the traditional ideal of rationality cannot be met, the Kuhnian and Arrovian views are diametrically opposed. Moreover, it makes a big difference which is right. If Kuhn is right that the problem is too many algorithms, two options suggest themselves. First, we might seek further conditions that any acceptable algorithm must satisfy, to narrow down the pool. Second, we might try to liberalize the notion of rationality, and argue that two scientists could both count as rational despite employing different algorithms for theory choice.[9] But if the Arrovian view is right, then neither of these options holds any promise, and our epistemological predicament is correspondingly more serious. Put differently, Kuhn makes rational theory choice look difficult, at least if we cleave to a certain conception of rationality, but Arrow makes it look outright impossible.

5. Possible escape routes

Since the Arrow-style impossibility result threatens the rationality of theory choice, and thus of science, it would be nice if there were a way out. Various possibilities suggest themselves. One is simply that many real cases of theory choice are binary, that is, involve just two alternatives. It is striking that Kuhn's own examples tend to be binary—geocentrism versus heliocentrism, oxygen versus phlogiston, relativity versus classical mechanics. And of course Arrow's impossibility result only holds if there are three or more alternatives. With just two alternatives to choose between,

[9] This idea was common in the post-Kuhn literature; see for example McMullin 1993. It is suggested by certain remarks of Kuhn himself, in the Postscript to *Structure of Scientific Revolutions*, where he says that his aim is not to show the irrationality of science, but rather to arrive at a more realistic view of what scientific rationality involves.

numerous algorithms become possible that satisfy conditions **U**, **P**, **I** and **N**.[10] Does this reconcile Kuhn with Arrow?

I think the answer is 'no', though it is certainly interesting that the distinction between binary and non-binary choice becomes significant, once theory choice is formulated in the social choice framework. (By contrast, on most standard philosophical approaches to theory choice, such as Bayesianism, it is of no particular relevance whether the choice is binary or not.) Not all theory choice in science is binary, even if the large-scale paradigm shifts that Kuhn is interested in typically are. More mundane cases are often not. Think for example of climate change science, where researchers compare the merits of numerous models of climate change, not just two. More generally, in many branches of science a typical problem might involve choosing between three candidate explanations of an observed correlation between two variables *x* and *y*: (i) *x* causes *y*; (ii) *y* causes *x*; (iii) *x* and *y* are joint effects of a common cause. Or consider statistical estimation, where a researcher might want to estimate the value of a real-valued parameter in the unit interval; the alternatives that must be chosen between are uncountably many. So focusing exclusively on binary choice, as a way of trying to avoid the Arrovian predicament, is at odds with scientific practice.

Another possibility for reconciling Arrow with Kuhn is simply to reject one or more of Arrow's conditions. Kuhn was sceptical (in some moods) about the existence of 'trans-paradigmatic' criteria of rationality, which are universally binding on scientists across all eras; perhaps he would argue that an acceptable algorithm for rational theory choice need not respect Arrow's conditions? This option does not seem especially plausible, given the intuitiveness of those conditions, but one possible argument against condition **N** (non-dictatorship) is worth briefly discussing.

In the context of theory choice, condition **N** says that no criterion can be a dictator, that is, can be such that whenever *x* ranks above *y* by that criterion, then *x* ranks above *y* overall. However, a strong empiricist might well hold that the criterion of 'fit-with-the-data' *should* be a dictator. Empiricists in philosophy of science have long argued that criteria such as simplicity are of lesser importance than fit-with-the-data, and should only be invoked, if at all, where the data cannot decide between two

[10] For example, 'majority rule', which says that alternative *x* is preferred to *y* just if it is ranked higher by more individuals (or criteria) is one such algorithm. With three or more alternatives, majority rule may lead the social preference to be intransitive, so does not even meet the definition of a social choice rule.

theories. Similarly, in a discussion of Kuhn's five criteria of theory choice, McMullin 1993 argued that 'accuracy' held a special role, for it is an end in itself while the others are only valuable in so far as they are reliable indicators of accuracy. So from an empiricist perspective, not all of Kuhn's criteria are equal.

Importantly, dictatorship of 'accuracy' (or 'fit-with-the-data') need not mean that the other criteria play no role at all in theory choice. For recall the definition of a dictator: a criterion (or individual) whose *strict* preference for *x* over *y* always leads *x* to be ranked higher than *y* overall. By contrast, a *strong* dictator is a criterion (or individual) whose preference for *x* over *y*, strict or weak, always becomes the overall preference. A strong dictatorship of 'fit with the data' would be an extreme form of empiricism—that refused to invoke extra-empirical criteria of theory choice even to break ties between pairs of theories that fit the data equally well. But an ordinary (not strong) dictatorship of 'fit with the data' could use criteria such as simplicity to break ties, that is, to settle cases where the dictator is indifferent. This is known as a 'serial' or 'lexicographic' dictatorship, and represents a more moderate form of empiricism.

Accepting a serial dictatorship of 'fit with the data' is in principle a way out of the impossibility result, since this theory choice rule does satisfy conditions **I**, **P**, and **U**. However, even if one accepts the underlying empiricist motivation, there are two problems with this solution. Firstly, to make it work, a complete lexicographic hierarchy of *all* the criteria of theory choice must be established, that is, a specification of the order in which they should be applied, to break ties. If there are only two criteria, for example fit-with-the-data and simplicity, then this is not a problem, but if there are more than two, there is a problem. For it is quite unclear how the hierarchy should be generated. Should simplicity or scope be invoked first, when fit-with-the-data cannot separate a pair of theories? Secondly and more importantly, a dictatorship of 'fit with the data', even serial, seems unattractive when we take account of the fact that our data invariably contain 'noise'. If our data were noise-free, always preferring a theory that fitted the data better would make sense. But with noisy data, perfect fit is not always desirable, as emphasised in the model-selection literature (Forster and Sober 1994). This 'problem of over-fitting', as it is known, constitutes a strong reason not to relax condition **N** in the manner mooted above, even if we are empiricists.

Finally, two further possible escape routes, well-known among social choice theorists, are worth briefly noting, though neither holds much promise.[11] The first is to modify the goal. Instead of trying to rank-order the alternatives, as in Arrow's formulation of the problem, suppose we instead try to pick the best. More precisely, we seek a 'choice function' which tells us, for any subset of the alternatives, which is (or are) the best. (In the scientific case this has some plausibility, as the problem of theory choice is often formulated as the problem of which theory to 'accept'.) This is a weaker goal than Arrow's, since a social preference relation entails the existence of a choice function but not vice-versa, thus holding some promise of an escape from the impossibility result. However, it turns out that if the choice function is required to satisfy certain quite reasonable properties, then analogues of Arrow's impossibility result re-emerge (Austen-Smith and Banks 1999). This escape route is thus thought unpromising by most social choice theorists, and seems equally unpromising as applied to theory choice.

The second option is domain restriction, i.e. dropping condition **U**. It is well-known that with a restricted domain, there may exist social choice rules that satisfy conditions **P**, **I**, and **N**. For certain applications of the social choice apparatus, 'natural' domain restrictions suggest themselves, though not for others. As noted in section 4, in the theory choice case, a natural domain restriction would apply if two of the criteria of theory choice exhibit an intrinsic trade-off (or correlation), for example, if a gain in simplicity always means a loss of accuracy. Then, certain profiles would be impossible, and could be legitimately excluded from the domain of the theory choice rule. However, that such trade-offs *always* exist does not seem very plausible; and anyway there is no guarantee that the resulting domain restriction would be of the right sort to alleviate the Arrovian impossibility.

6. Sen's 'informational basis' approach

I turn now to what is arguably the most attractive 'escape route' from Arrow, namely Amartya Sen's idea of using an 'enriched informational basis' (Sen 1970, 1977,

[11] A third possible escape route is to relax condition **I**, a move defended by some authors, e.g. Fleurbaey and Maniquet (2008), in the context of choosing allocations for an economy. However it seems unlikely that the rationale for relaxing **I**, in such contexts, transposes to the context of theory choice discussed here.

1986).[12] Sen observes that the information Arrow uses as input to his social choice rule, namely a profile of individual preference orders, is quite meagre. This is for two reasons. Firstly, preference orders are 'purely ordinal'—they contain no information about *intensity* of preference. If an individual prefers *x* to *y* to *z*, this tells us nothing about whether their preference for *x* over *y* is greater or less than their preference for *y* over *z*. Secondly, preference orders do not permit interpersonal comparisons. From a profile of individual preference orders, statements such as 'in alternative *x*, individual 1 is better off than individual 2' cannot be deduced.

To remedy these problems, Sen suggests that we start not with a profile of preference orders, but rather of *utility functions*, one for each individual in society. An individual's utility function assigns a real number to each alternative, which reflects how much utility that alternative would bring them. Let u_i denote the utility function of the i^{th} individual; let $<u_1, \ldots, u_n>$ denote a profile of utility functions. An individual's utility function is required to *represent* their preference order, in the sense that xR_iy iff $u_i(x) \geq u_i(y)$, for all alternatives *x* and *y*. (Recall that 'xR_iy' means that individual i weakly prefers *x* to *y*.) Note that if u_i represents R_i, then any increasing transformation of u_i will also represent R_i. Thus there is a many-one relation between utility functions and the preference orders that they represent.

Next, Sen introduces the concept of a *social welfare functional* (SWFL). This is a function that takes as input a profile of utility functions, and yields as output a social ranking of the alternatives. A SWFL is analogous to an Arrovian social choice rule, in that both yield the same output; however, the former takes a profile of utility functions, rather than preference orders, as input. Potentially, this allows more information to be taken into account.

Analogues of Arrow's four conditions can now be imposed on the SWFL. The analogue of **U** says that the domain of the SWFL is the set of all possible profiles of utility functions, that is, individuals can have whatever utility functions they please. The analogue of **P** says that if everyone gets more utility in alternative *x* than in *y*, then *x* is socially preferred to *y*. The analogue of **N** says that there can be no individual such that whenever they get more utility from *x* than *y*, then *x* is socially preferred to *y*. The analogue of **I**, known as 'independence of irrelevant utilities', says that the social preference between *x* and *y* must depend only on individuals' utilities in

[12] Sen's approach has been further developed by numerous workers; for good overviews see Gaertner 2006, Roemer 1997, and Bossert and Weymark 2004.

x and *y*. These conditions on the SWFL will be denoted **U′**, **P′**, **I′** and **N′**; they are motivated by arguments similar to those that motivate the Arrovian originals.

One might think that an analogue of Arrow's impossibility result will now apply, that is, that no SWFL can satisfy conditions **U′**, **P′**, **I′** and **N′**. However, this is not correct. Arrow's impossibility result *can* be derived in Sen's framework, but it requires an additional condition, capturing the fact that Arrow uses purely ordinal, non-interpersonally comparable information. To see how this informational assumption can be captured, consider a profile of utility functions $<u_1, \dots, u_n>$. Now suppose each of the n individuals applies an increasing transformation to their utility function, yielding a new profile $<v_1, \dots, v_n>$. (Different individuals may apply different transformations.) On Arrow's assumption, the two profiles contain exactly the same information – since the transformed utility functions represent the very same preferences. So Arrow will argue that the SWFL should yield the same social ranking when applied to the two profiles. This condition is called 'invariance with respect to ordinal, non-comparable information' or **ONC**. Arrow's theorem can now be stated in Sen's framework: for three or more social alternatives, no SWFL can satisfy conditions **ONC**, **U′**, **P′**, **I′** and **N′**.

If the **ONC** condition is imposed on the SWFL, this implies that interpersonal comparison of utility is deemed impossible (or meaningless). To see why, suppose that in profile $<u_1, \dots, u_n>$, individual 1 gets more utility than individual 2 in a given alternative *x*, that is, $u_1(x) > u_2(x)$. But this inequality is not necessarily preserved, if the individuals apply different positive transformations to their utility functions. So in the transformed profile $<v_1, \dots, v_n>$, it need not be true that $v_1(x) > v_2(x)$. Therefore, if the two profiles are treated as informationally equivalent, as the **ONC** condition demands, it follows that interpersonal comparisons cannot be made.

The natural next step is to ask what happens if the **ONC** condition is relaxed. There are two ways it can be relaxed: (i) drop the assumption that utility is purely ordinal; (ii) permit interpersonal comparisons. To effect (i), we restrict the transformations that can be applied to a given utility function; to effect (ii), we cease to allow individuals to choose their own transformations independently of others. Let us take (i) first. Instead of ordinal utility, we might hold that utility is measured on a *cardinal* scale, so only positive linear transformations, of the form $v_i = au_i + b$, $a>0$,

are held to preserve information.[13] In effect, this means that an individual's utility function contains information about the intensity of their preferences, so utility differences become meaningful. Alternatively, we might hold that utility is measured on a *ratio* scale, so only transformations of the form $v_i = au_i$, $a>0$, are held to preserve information. This means that the utility scale has a natural zero point, so utility ratios become meaningful.[14] Finally, we might hold that utility is measured on an *absolute* scale, that is, only the identity transformation preserves information. This means that actual utility numbers are meaningful.

Once a scale for utility has been chosen—ordinal, cardinal, ratio or absolute—a decision about interpersonal comparability is necessary. If utility is *non-comparable*, then each individual can apply a transformation (from the permissible class) independently of others. If utility is *fully comparable*, then each individual must apply the same transformation. Depending on the utility scale, a form of partial comparability may also be possible. With cardinal utility, if utility is *unit comparable*, then individuals' positive linear transformations must all have the same slope, but can have different intercepts.

Numerous alternatives to Arrow's **ONC** condition are now possible. They include: cardinal-scale utility with no comparability (**CNC**); cardinal-scale utility with full comparability (**CFC**); ratio-scale utility with full comparability (**RFC**); ratio-scale utility with no comparability (**RNC**); and absolute-scale utility with full comparability (**AFC**). In effect, each of these conditions partitions the set of all profiles of utility functions into equivalence classes of 'informationally equivalent' profiles, and requires that the SWFL yield the same social ranking for all the profiles in a given equivalence class. **ONC** is the strongest condition—for the classes of profiles that it treats as informationally equivalent are very large, and thus the restriction on the SWFL considerable. By contrast, **AFC** is the weakest condition—it places each profile into a singleton class of its own, which implies no restriction on the SWFL. This illustrates a general moral: the richer the informational basis, that is, the finer the partition of the profiles into equivalence classes, the weaker the resulting condition on the SWFL.

Sen now asks: what happens if we retain the four Arrovian conditions **U′**, **P′**, **I′** and **N′**, but replace **ONC** with a weaker condition? Can the impossibility result be

[13] Temperature in celsius or farenheit is measured on a cardinal, or interval, scale.
[14] Length in centimetres is an example of a quantity measured on a ratio scale.

avoided? The answer is that impossibility *can* be avoided, but only if some interpersonal comparison is allowed. Replacing **ONC** with **CNC**, that is, moving from ordinal to cardinal utility, is no help on its own. The same is true of moving to ratio-scaled utility (**RNC)**. However, if **ONC** is replaced with **CFC**, **RFC**, or **AFC**, then Arrovian impossibility is avoided. There do exist social welfare functionals that satisfy Arrow's four conditions, plus one of these alternatives to **ONC**.

(The case of ratio-scale non-comparability (**RNC**) merits further discussion, for a reason that will become clear. Although replacing **ONC** with **RNC** does not avoid Arrovian impossibility, *it does do if all utilities are required to be non-negative* (Tsui and Weymark 1997). If all utilities are non-negative, there *do* exist social welfare functionals that satisfy **RNC** and Arrow's four conditions. Also, note that **RNC**, despite its name, does permit a limited sort of interpersonal utility comparison.[15] With **RNC**, *percentage increases in utility* can be meaningfully compared, that is, statements such as 'in moving from alternative *x* to *y*, individual 1's percentage gain is greater than individual 2's', are meaningful (Fishburn 1987). It is easy to verify that the truth-value of this statement will be unaltered, if the two individuals apply different ratio-scale transforms to their utility functions.)

Sen's analysis raises two issues. First, how can interpersonal comparisons of utility be made? Second, once such comparisons are allowed, how large is the class of SWFLs that satisfy the Arrow conditions? Can further conditions be found that narrow down the permissible SWFLs to a single one? There is an extensive literature on both these points, but space does not permit them to be explored here.[16] For the moment, the point to note is just this. Sen's work demonstrates clearly that Arrow's impossibility result is in large part a consequence of the impoverished information he feeds into his social choice rule. Enriching the informational basis, while retaining Arrow's four conditions—now understood as conditions on the social welfare functional, rather than the social choice rule—is sufficient to avoid the impossibility.[17] In short, given enough information, reasonable social choices can be made.

---

[15] For this reason, some authors prefer the label **RSM** (ratio-scale measurability) for what I am calling **RNC**.

[16] On the first issue see the papers in Elster and Roemer 1991; on the latter, see Gaertner 2006, Roemer 1997 or Bossert and Weymark 2004.

[17] In a way, Sen's approach involves a rejection of Arrow's condition **I** (independence of irrelevant alternatives). For Arrow's condition **I** is equivalent to the conjunction of **ONC** and condition **I′**

7. Theory choice: the informational basis

Let us return to theory choice and apply the morals of the previous section. Recall that we defined a theory choice rule, by direct analogy with Arrow's social choice rule, as a function that takes as input a profile of weak orders, one for each criterion of theory choice, and outputs an 'overall ranking' of the alternative theories. Just as Sen replaced Arrow's social choice rule with a social welfare functional, so we need to replace our theory choice rule with a 'theory choice functional'. So instead of starting with a profile of 'preference orders', one for each criterion of theory choice, we start with a profile of 'utility functions', that is, real-valued representations of those orders. In principle, this allows an enrichment of the informational basis.[18]

The natural next question is: which profiles should be treated as informationally equivalent, that is, which invariance condition should be imposed on the theory choice functional? To address this question, we need to consider both measurement scales and 'inter-criterion' comparability. Let us take them in turn. For some criteria of theory choice, an ordinal scale might be appropriate. Kuhn's criterion of 'fruitfulness' is an example. Conceivably, one could order a set of theories by how fruitful they are, but it is hard to believe that *differences* in fruitfulness can be compared; a statement such as 'the difference in fruitfulness between T_1 and T_2 exceeds the difference between T_2 and T_3' hardly seems meaningful. If this is right, then the real-valued 'utility' function that represents the fruitfulness preference order is merely ordinal—any increasing transformation can be applied to it without loss of information.

However for other criteria of theory choice, we can go beyond ordinal measurement, at least in certain contexts. Take for example fit-with-the-data (or 'accuracy'), and suppose that the context is linear regression analysis. The usual measure of how well a hypothesis fits the data in linear regression is its 'sum of squares' (SOS) score.[19] The appropriate type of measurement scale for SOS scores depends on the dependent variable in the regression model. If for example that variable is length, which is a ratio-scale measurable quantity, then the SOS scores will

---

(independence of irrelevant utilities). Retaining **I′** while rejecting **ONC** thus abandons the letter of Arrow's original independence condition, while retaining its spirit.

[18] This application of Sen's framework to the problem of theory choice illustrates a point made by Kelsey 1987, namely that the functions used in an SWFL need not necessarily be interpreted as utility functions.

[19] The SOS of a hypothesis is the sum of the squared distance of each data point from the hypothesis's prediction.

also be ratio-scale measurable. Therefore, the real-valued 'utility' function that represents the 'fit-with-the-data' preference order will be ratio-scaled, thus multiplication by a positive constant is the only information-preserving transformation. Statements such as 'T_1 fits the data three times as well as T_2' will be meaningful.

To take another example of how we can often go beyond ordinal information, consider simplicity. In certain contexts, such as statistical model selection, the simplicity of a hypothesis is taken to be the number of free parameters it contains. Thus for example, the hypothesis '$y = ax + b$' is simpler than '$y = ax^2 + bx + c$' because the former contains two free parameters, the latter three. So in this case, simplicity is measured on an absolute scale—the actual numbers are meaningful, so only the identity transformation preserves information. Similarly, in a Bayesian context, a prior probability distribution over a set of hypotheses is a case of absolute measurability—the actual numbers assigned are meaningful. So in both these cases, we have much more than ordinal information.

This suggests that the question of what measurement scales are appropriate, for criteria of theory choice, does not have a simple answer. Different scales may be appropriate for different criteria, and may depend on the inferential techniques that we are using. It may be that for the 'large scale' theory choices that Kuhn was interested in, ordinal comparisons are all that can be achieved. But it seems clear that in other, more humdrum cases, particularly where the problem may be formulated statistically, we may have much more than ordinal information at our disposal.

Finally, note that the situation for theory choice is more complicated than for social choice. In social choice, one normally assumes a single type of measurement scale for all utility functions. It would make little sense to suggest that individual 1's utility function was ordinal, individual 2's cardinal. But the analogous situation for theory choice makes good sense. It might well be that fruitfulness, for example, is merely ordinal but that fit-with-the-data is ratio-scale measurable.

What about inter-criterion comparisons, the analogue of interpersonal comparisons? One might think that such comparisons are unlikely. Take for example the statement: 'the difference in simplicity between T_1 and T_2 exceeds the difference in accuracy between T_3 and T_4'. It is hard to see what the basis for such a judgement might be. It is harder still to see how comparisons of levels, rather than differences, could be made—this would permit statements such as 'the accuracy of T_1 is less than

the simplicity of T_2', which sound even odder. Since inter-criterion comparability is needed to avoid the impossibility result, as we know, the prospects for escaping the Arrovian predicament by enriching the informational basis of theory choice may seem dim.

However this is overly pessimistic, for two reasons. Firstly, note that if all criteria are absolutely measurable, then interpersonal comparability follows immediately. If the 'utility' functions that represent the simplicity and accuracy orderings cannot be transformed without loss of information, then statements such as 'the accuracy of T_1 is less than the simplicity of T_2' automatically become meaningful. (Crucially, 'meaningful' here has a technical sense, i.e. invariance under the permissible transformations; it does not mean that there would be any particular purpose in uttering the statement in question.)[20] Since, as argued above, absolute measurability may be appropriate for some criteria of theory choice in some contexts, inter-criterion comparability should not be dismissed out of hand.

Secondly, recall the discussion of ratio-scale measurability in section 6. If the criteria of theory choice are each measured on their own ratio-scale (i.e. **RNC**), then this: (i) permits a limited form of inter-criterion comparability, and (ii) avoids Arrovian impossibility so long as all 'utilities' are non-negative. Ratio-scale measurability is fairly plausible in certain inferential contexts. Consider 'scope', for example. If differences in scope can be compared, and if in addition there is a natural zero point, that is, it makes sense to talk about a theory with zero scope, then scope is ratio-scale measurable. This does not seem altogether implausible, for some criteria in some inferential contexts. If both scope and accuracy (say) are ratio-scale measurable, each with their own scale, then this permits a limited form of inter-criterion comparability: percentage increases in scope may be compared with percentage increases in accuracy. (So statements such as 'T_1 has 10% less scope than T_2, but is 15% more accurate' can be made.) As regards point (ii), the restriction to non-negative 'utilities' seems unproblematic; if 'scope' has a natural zero point, why demand that the theory choice functional be able to deal with profiles in which some theories (*per impossible*) have negative scope? So there is a potential escape route from Arrow here too.

[20] It is also important to see that inter-criterion comparability does not require that the two criteria be measured in the same units, for the comparison in question is a comparison of real numbers, not of the quantities that they represent.

To sum up, Sen's work, transposed to the theory choice case, tells us that there do exist theory choice functionals that satisfy Arrow's four conditions, so long as the **ONC** condition is replaced in favour of one that permits inter-criterion comparison. This prompts the question of what replacement of **ONC** (if any) is appropriate, that is, which profiles of 'utility' functions should be treated as informationally equivalent. There is no simple answer to this question. However, in some cases absolute measurement will be appropriate, implying that **ONC** should be replaced with **AFC**; this permits the impossibility result to be avoided. In other cases ratio-scale measurement will be appropriate, which also permits the impossibility result to be avoided. The general moral is that enriching the informational basis of theory choice *does* permit an escape from Arrow; though which enrichments are defensible must be answered on a case-by-case basis.

Where does this leave us vis-à-vis Kuhn's 'no algorithm' thesis? If we can escape the Arrovian predicament by enlarging the informational basis, as described above, we will end up with *many* theory choice functionals that meet our reasonableness constraints. For replacing **ONC** with an alternative condition (such as **AFC**), while retaining Arrow's four conditions, does not narrow down the class of permissible theory choice functionals to a single one. So we escape Arrow's predicament only to enter Kuhn's: many acceptable algorithms, and no way to select between them. To escape both predicaments, we need reasonableness conditions that are satisfied by exactly one algorithm. In the social choice literature, researchers have managed to identify conditions that uniquely pick out particular social welfare functionals, such as the utilitarian SWFL, Rawlsian maximin, and others; but it is doubtful whether the analogues of these conditions, transposed to the theory choice case, would be defensible. (By contrast, the analogues of Arrow's conditions are certainly defensible.) Therefore in the theory choice case, escaping Arrovian impossibility by enriching the informational basis seems to lead us straight to Kuhn's 'no algorithm' thesis.

8. Illustrations: Bayesianism and statistical model selection

The previous section's main conclusion—that Sen's escape route from Arrow does apply to theory choice—can be illustrated by considering the orthodox Bayesian approach to scientific inference. Suppose we have a body of evidence E, and five rival hypotheses $\{T_1, \ldots, T_5\}$ that are pair-wise exclusive. On the Bayesian view, we use

two criteria to choose between the hypotheses: prior probability $P(T_i)$, and likelihood, $P(E/T_i)$. (Likelihood can be thought of as a measure of the 'fit' between evidence and hypothesis, prior probability a measure of the antecedent plausibility of a hypothesis, before the evidence.) Clearly, there are many possible ways of combining these two criteria into a decision rule; but Bayesians argue that the right way to do it is to multiply the prior by the likelihood, that is, to consider the quantity $[P(T_i) \times P(E/T_i)]$. The theory with the highest value of this quantity is the most deserving of our credence, according to Bayesians; and more generally, this quantity can be used to generate an overall ranking of the theories, from best to worst.

Bayesians have some sophisticated arguments for why this is the right way to combine the two criteria, but they need not detain us. For the moment, we want to relate Bayesianism to our foregoing discussion. To do this, simply think of $P(T_i)$ and $P(E/T_i)$ as 'utility' functions, both of which assign a real number to each of the five theories. The ordered pair $< P(T_i), P(E/T_i)>$ is then a *profile*, corresponding to a profile of utility functions in a two-person society, in Sen's framework. Now consider the function which maps the set of profiles onto the ranking generated by the quantity $[P(T_i) \times P(E/T_i)]$. This corresponds to a social welfare functional in Sen's framework, or what we called a 'theory choice functional' in section 7. The functional takes as input the prior probability and likelihood of each theory, and yields as output a ranking of the theories, from best to worst. Let us call this the 'Bayesian theory choice functional', or BCF.

Now let us ask: does the BCF satisfy the four Arrovian conditions? Consider firstly condition **P′**, weak Pareto. It easy to see that **P′** is satisfied: if theory T_1 has a higher prior *and* a higher likelihood than theory T_2, that is, $P(T_1) > P(T_2)$ and $P(E/T_1) > P(E/T_2)$, then T_1 will obviously be ranked higher than T_2 by the BCF. Condition **I′**, the independence of irrelevant 'utilities' condition, is also satisfied: whether T_1 or T_2 is ranked higher by the BCF is entirely determined by the priors and likelihoods of those two theories; no other information is relevant. Finally, condition **N′**, non-dictatorship, is also satisfied. Neither criterion (prior or likelihood) is able to dictate over the other—it is not true that if T_1 has a higher prior than T_2 then it must be ranked higher, and similarly for likelihood.

What about condition **U′**, unrestricted domain? This says that the domain of the theory choice functional must be the set of all possible profiles, that is, pairs of real-valued functions. Clearly the BCF does not satisfy this condition, for both of the

functions that we feed into it, $P(T_i)$ and $P(E/T_i)$, can only take on values in the unit interval [0,1]; moreover, it is required that $\sum P(T_i) \leq 1$. Thus there are two restrictions on the permissible values of the functions we feed into the Bayesian theory choice functional. So condition **U′** is not satisfied, whereas conditions **P′**, **I′** and **N′** are.

Mindful of Arrow's theorem, one might think that it is *because* the BCF has a restricted domain, so violates condition **U′**, that it can satisfy **P′**, **I′** and **N′**. But this is not correct. Recall that within Sen's framework, the derivation of Arrow's result requires the **ONC** condition, in addition to conditions **U′**, **P′**, **I′** and **N′.** The **ONC** condition says that any two profiles are informationally equivalent, hence should be mapped to the same ranking, if one is derivable from the other by applying increasing transformations to the functions in the profile. But the BCF does not satisfy this condition; on the contrary, it is quite possible to have two profiles $< P(T_i), P(E/T_i)>$ and $< Q(T_i), Q(E/T_i)>$, where the two prior functions $P(T_i)$ and $Q(T_i)$ rank the theories identically, and the two likelihood functions $P(E/T_i)$ and $Q(E/T_i)$ also rank them identically, and yet the overall rankings, generated by the BCF, are different in the two cases.

This prompts the question: which measurability/comparability assumption *is* appropriate for the Bayesian theory choice functional? Since probabilities are measured on an absolute scale, the answer is clear: absolute full comparability (**AFC**). Given a profile $< P(T_i), P(E/T_i)>$, applying any transformation to it other than the identity transformation will alter its informational content—for the actual probability numbers are meaningful.[21] So the prior probability function contains much more than the merely ordinal information that gives rise to Arrovian impossibility; the same is true of the likelihood function.

Therefore, the Bayesian theory choice functional violates two of the conditions that are required to generate Arrow's impossibility result in Sen's framework—**U′** and **ONC**. The fact that the BCF has a restricted domain is not crucial—what matters is the fact that it uses more than ordinal non-comparable information. For even with the domain restriction appropriate to the BCF, if the **ONC** condition were imposed, then the **P′**, **I′** and **N′** conditions would be jointly unsatisfiable (see Appendix for proof.) The fact that the BCF satisfies the latter three

[21] Probabilities are measured on an absolute scale *modulo* the convention that the sure event has probability 1. If this convention were relaxed, i.e. if we chose some other positive number for the probability of the sure event, then probability would become ratio-scale measurable.

conditions is therefore attributable to the richness of the information fed into it, rather than to its restricted domain.

This illustrates how Sen's escape route from Arrow does apply to theory choice. The Bayesian theory choice functional constitutes a kind of algorithm for theory choice, and does satisfy the Arrovian conditions (other than **U′**); this is possible because the appropriate measurability/comparability assumption is **AFC**, rather than **ONC**. The BCF is by no means the *only* theory choice functional that satisfies **AFC** and the Arrovian conditions; so there is potential for a Kuhnian 'no unique algorithm' argument. However, such an argument would have to counter the Bayesians' argument for why their theory choice functional is the 'correct' one. This important issue cannot be addressed here. My aim has only been to illustrate how in theory choice, enriching the informational basis permits an escape from Arrovian impossibility, just as it does in social choice.

A second illustration of this point is provided by a quite different approach to scientific inference, namely statistical model selection. This approach is a sophisticated variant of what philosophers call the 'curve fitting' problem, or inferring the functional relation between two variables from finite data. However, unlike in more simplified discussions of curve-fitting, it is assumed that the data are 'noisy', that is, the data points are affected by measurement error. The aim is to choose between alternative hypotheses about the functional relation between two variables x and y, from the noisy data. Two criteria are used: simplicity and fit-with-the-data. The two will often exhibit a trade-off: improving fit means sacrificing simplicity. Clearly, we have here the ingredients for a Kuhnian 'no unique algorithm' claim – there are many conceivable ways of combining fit and simplicity into a single decision rule.

In a typical model selection problem, we start with a number of *families* of hypotheses. For example, one family consists of all hypotheses of the form $y = a + bx$, $a, b \in \Re$; let us call this family **LIN**. Another family, **PAR**, consists of all hypotheses of the form $y = a + bx + cx^2$; and a third, **EXP**, consists of all hypotheses of the form $y = a^x$. The real numbers a, b and c are called *free parameters*; and the simplicity of a hypothesis is defined as the number of free parameters in its family. Thus each hypothesis in **LIN** is simpler than each in **PAR**. Next, one finds the best-fitting hypothesis in each family, denoted L(**LIN**), L(**PAR**) and L(**EXP**) respectively; the criterion of best-fit is highest likelihood, where 'likelihood' has its customary

statistical meaning (roughly, the probability of observing the actual data, if the hypothesis were true.)

In the statistical literature, there exist various suggestions for how to combine simplicity and fit into a decision rule (Forster 2001); such a rule would allow us to choose between the three hypotheses *L*(**LIN**), *L*(**PAR**) and *L*(**EXP**), in the above example. One of the best-known is the *Akaike criterion*, which says that that we should choose the hypothesis with the highest *Akaike score*; the Akaike score of a hypothesis H is defined as [log-likelihood H − *k*], where *k* is the number of free parameters of the family to which the hypothesis belongs. Therefore, the better a hypothesis fits the data the higher its Akaike score; however, there is a penalty for complexity, that is, for having lots of free parameters. Thus Akaike's criterion combines simplicity and fit-with-the-data into a single algorithm for hypothesis choice, which ranks the hypotheses from best to worst. Obviously, there are many other conceivable algorithms, but there exist sophisticated arguments for why Akaike's is the 'correct' one.

To relate statistical model selection to social choice, think of log-likelihood H_i and k_i as 'utility functions', each of which assigns a number to each hypothesis reflecting, respectively, its fit and its simplicity. Then consider the set of all ordered pairs (profiles) of the form <log-likelihood H_i, k_i>—these are all the possible combinations of fit with simplicity. Then consider the function from this set to the ranking of hypotheses generated by the Akaike score. Formally, this function corresponds to the social choice functional of section 6 and the theory choice functional of section 7. Let us call it the 'Akaike choice functional'.

Now we can ask: does the Akaike choice functional satisfy the Arrovian conditions? It is easy to see that conditions **P′**, **I′** and **N′** are all satisfied, by an argument parallel to the one given above for the Bayesian choice functional. But condition **U′** (unrestricted domain) is not satisfied—for the log-likelihood function only takes negative values and the free parameter function only takes positive integer values, which implies a domain restriction. What about the **ONC** (ordinal non-comparability) assumption? This is not satisfied either. It is not the case that if one profile can be got from another by applying increasing transformations to the two 'utility' functions, that the Akaike choice functional will necessarily map them to the same ranking. Again, the appropriate assumption, in lieu of **ONC**, is absolute full comparability (**AFC**); since the actual numbers assigned by the two functions are

meaningful, any transformations will change their informational content. As with the Bayesian choice functional, it is because the Akaike choice functional does not satisfy **ONC**, rather than because of its restricted domain, that it is able to satisfy conditions **P′**, **I′** and **N′**.[22]

Again, the point of this is not to defend the Akaike criterion in particular, but rather to illustrate how Sen's moral—that enriching the informational basis can avoid Arrovian impossibility—applies to theory choice. The information we feed as input into the Akaike choice functional is far more than merely ordinal, which explains why it satisfies the Arrovian conditions (other than **U′**). Of course there are many alternative functionals, besides Akaike's, that will also satisfy those conditions, so again, escaping the Arrovian predicament may land us in the Kuhnian one. However, there are also arguments for why the Akaike criterion is the uniquely 'correct' one; so it may be that both predicaments can be avoided. A proper assessment of this issue cannot be undertaken here.

9. Conclusion

Although Kuhn's 'no algorithm' thesis is quite widely accepted in philosophy of science, there have been few attempts to subject it to serious scrutiny. To remedy this situation, I have used the machinery of social choice theory, and tried to relate Kuhn's thesis to Arrow's famous impossibility theorem. Though superficially similar to Kuhn's, Arrow's conclusion that there is 'no algorithm' for social choice is in fact quite different. For Kuhn's claim is that there are many algorithms, all equally acceptable, while Arrow's claim is that *no* algorithms meet minimum standards of acceptability.

By identifying Kuhn's five criteria with Arrow's individuals, the theory choice problem was seen to have the same structure as a standard social choice problem. Moreover, Arrow's four conditions seem as defensible for theory choice as they are for social choice, which raises the spectre of an Arrovian impossibility result for theory choice. Such a result would constitute a refutation of Kuhn's thesis, but would also pose a threat for the rationality of science; a threat that if anything is more worrying than that posed by Kuhn.

[22] An argument parallel to the one given in the Appendix shows that with the domain restriction appropriate to the Akaike choice functional, conditions **P′**, **I′** and **N′** are jointly unsatisfiable if **ONC** is assumed. Thus it is the violation of **ONC**, not the domain restriction, that permits the satisfaction of conditions **P′**, **I′** and **N′.**

A number of possible 'ways out' of the impossibility, while remaining within Arrow's framework, were canvassed; these included confining attention to binary theory choices and accepting a serial dictatorship of 'accuracy'. More promising was Sen's idea of moving to a different framework by enriching the informational basis, that is, going beyond the ordinal, non-comparable information that Arrow starts with. Applying Sen's idea to theory choice raised difficult questions about the appropriate measurability/comparability assumptions; however, we showed by example that two well-known approaches to theory choice, Bayesianism and statistical model selection, avoid Arrovian impossibility precisely by incorporating more than ordinal, non-comparable information.

Finally, what then of Kuhn's 'no algorithm' thesis? Is it correct? No simple answer to this question emerges from the foregoing analysis. It may be that we can avoid Arrovian impossibility only by opening the door to many different algorithms, thus vindicating Kuhn; but we should not rule out the possibility of finding additional rationality constraints which considerably narrow down the acceptable ones, possibly even to uniqueness. Although my analysis does not provide a definitive resolution, I hope to have identified the *sorts* of consideration that are relevant to determining whether Kuhn's thesis is correct.[23]

Department of Philosophy, SAMIR OKASHA
University of Bristol,
9 Woodland Road,
Bristol, BS8 1TB
Samir.Okasha@bristol.ac.uk

[23] Thanks to audiences at Bristol, Cambridge, Konstanz, Madrid and Cardiff where versions of this paper were presented; to Elliott Sober, Marcel Weber, Hannes Leitgeb, Armin Schulz and Kit Patrick for comments and discussion; and in particular to an anonymous referee for *Mind* for extremely detailed comments. This work was supported by AHRC grant AH/F017502/1.

Appendix

Consider a finite set X of pair-wise exclusive theories $\{T_1, \ldots, T_n\}$. We assume X is a partition of logical space.[24] Let Y be the set of all orderings of X.

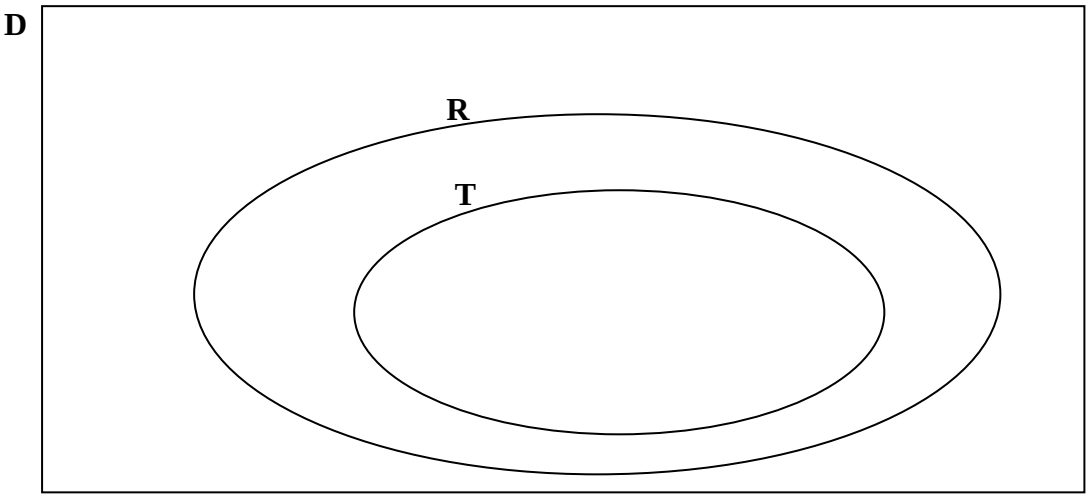
There are two 'individuals', each with a real-valued 'utility' function over X, denoted u_1 and u_2 respectively. A profile (ordered pair) of utility functions, one for each individual, is denoted $<u_1, u_2>$.

Let D be the set of all possible profiles.

Let $R \subset D$ be the set of all profiles satisfying the condition: for all individuals i, and all $x \in X$, $0 \leq u_i(x) \leq 1$, i.e. both utility functions can only take values in [0,1].

Let $T \subset R$ be the set of all profiles satisfying the additional condition: $\sum u_1(T_i) = 1$, i.e. the utilities that individual 1 assigns to the theories sum to 1.

So we have $T \subset R \subset D$ (see diagram).



By Arrow's theorem (in Sen's framework), there exists no function f: D→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

We wish to show that there is no function f′: T→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

We show firstly that there is no function g: R→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

[24] Taking X to be a partition makes the second half of the proof easier, as it allows us to assume that $\sum u_1(T_i) = 1$ as opposed to merely $\sum u_1(T_i) \leq 1$, but could easily be relaxed.

1. Suppose for *reductio* that there exists a function g: R→Y satisfying conditions **P′**, **I′**, **N′** and **ONC**.

Let S ⊂ R be the set of all profiles satisfying the condition: for all individuals i, and all $x \in X$, $0 < u_i(x) < 1$.

Since g satisfies **P′**, **I′**, **N′** and **ONC**, and since g has domain R and S ⊂ R, then there exists a function g′: S→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

Let t: ℜ→(0,1) be any increasing function mapping the real number line onto the open unit interval, e.g. $t(x) = [x\text{-}2]^{-1}$ if $x < 0$, $t(x) = [x+1]/[x+2]$ if $x \geq 0$.

Let h be the result of applying the transformation $u_i(x) \rightarrow t[u_i(x)]$ to each individual's utility function in each profile $\langle u_1, u_2 \rangle$ in D. Therefore h: D → S maps each profile in D onto a profile in S.

The function g′ o h ('g′ after h') then has domain D and range Y.

Since g′ satisfies conditions **P′**, **I′**, **N′** and **ONC**, and since t is increasing, it follows that g′ o h satisfies conditions **P′**, **I′**, **N′** and **ONC** too. But by Arrow's theorem no function with domain D satisfies these conditions.

Therefore there is no function g: R→Y satisfying conditions **P′**, **I′**, **N′** and **ONC.**

2. Suppose for *reductio* that there exists a function f′: T→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

Let r: [0,1] →[0,1] be a function mapping individual 1's utility function $u_1(T_i)$ onto $[u_1(T_i) / [(u_1(T_1) + u_1(T_2) +, \ldots , + u_1(T_n)]]$. (Applying r to individual 1's utility function has the effect of normalizing it, to ensure that $\sum u_1(T_i) = 1$.)

Let h′ be the result of applying the transformation $\langle u_1, u_2 \rangle \rightarrow \langle r(u_1), u_2 \rangle$ to each profile in R. Therefore h′: R→T maps each profile in R onto a profile in T.

The function f′ o h′ ('f′ after h'') then has domain R and range Y.

Since f′ satisfies conditions **P′**, **N′** and **ONC**, and since r is increasing, it follows that f′ o h′ satisfies conditions **P′**, **N′** and **ONC** too.

To show that f′ o h′ satisfies condition **I′**, consider two profiles $\langle u_1, u_2 \rangle$ and $\langle v_1, v_2 \rangle$ in R that coincide over the theories T_1 and T_2, i.e. $u_1(T_1) = v_1(T_1)$, $u_2(T_1) = v_2(T_1)$ and $u_1(T_2) = v_1(T_2)$, $u_2(T_2) = v_2(T_2)$.

[Note: it does *not* follow that $\langle r(u_1), u_2 \rangle$ and $\langle r(v_1), v_2 \rangle$ must also coincide over T_1 and T_2.]

Choose any profile $\langle r'(u_1), u_2\rangle$ in T such that:

(i) $r'(u_1)$ and $r(u_1)$ order the theories in the same way; and

(ii) $r'(u_1)(T_1) = r(v_1)(T_1)$, and $r'(u_1)(T_2) = r(v_1)(T_2)$

(There must be such a profile, since $r(u_1)$ and $r(v_1)$ order T_1 and T_2 in the same way; for all other theories T_i, choose $r'(u_1)(T_i)$ so as to produce the same order as $r(u_1)$.)

Now, the profiles $\langle r'(u_1), u_2\rangle$ and $\langle r(v_1), v_2\rangle$ coincide over T_1 and T_2, so f′ must map them onto orderings of the set X in which T_1 and T_2 are ranked identically, by condition **I′**.

Since $r'(u_1)$ and $r(u_1)$ order the theories in the same way, f′ must map $\langle r'(u_1), u_2\rangle$ and $\langle r(u_1), u_2\rangle$ onto the same ordering of the set X, by condition **ONC**.

Therefore, f′ must map $\langle r(u_1), u_2\rangle$ and $\langle r(v_1), v_2\rangle$ onto orderings of the set X in which T_1 and T_2 are ranked identically.

Therefore, f′ o h′ maps $\langle u_1, u_2\rangle$ and $\langle v_1, v_2\rangle$ onto orderings of the set X in which T_1 and T_2 are ranked identically, i.e. f′ o h′ satisfies condition **I′.**

So f′ o h′ has domain R and satisfies conditions **P′**, **I′**, **N′** and **ONC**. However, by part 1 above, no function with domain R satisfies those conditions.

Therefore, there is no function f′: T→Y satisfying the conditions **P′**, **I′**, **N′** and **ONC**.

References

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

Arrow, K. and M. Instiligator (eds) 1986: *Handbook of Mathematical Economics*. North-Holland: Elsevier.

Austen-Smith, David and Jeffrey S. Banks 1999: *Positive Political Theory 1: Collective Preference*. Ann Arbor: University of Michigan Press.

Barbera, S., P. Hammond, and C. Seidl (eds) 2004: *Handbook of Utility Theory*, volume 2. Dordrecht: Kluwer.

Bossert, Walter and John A. Weymark 2004: 'Utility in Social Choice'. In Barbera, Hammond, and Seidl 2004, pp. 1099−177.

Earman, John 1992: *Bayes or Bust?* Cambridge, MA: MIT Press.

Elster, John and John E. Roemer (eds) 1991: *Interpersonal Comparisons of Well-Being*. Cambridge: Cambridge University Press.

Fishburn, Peter C. 1987: *Interprofile Conditions and Impossibility*. London: Routledge.

Fleurbaey, Marc and Francois Maniquet 2008: 'Utilitarianism versus Fairness in −Welfare Economics'. In Fleurbaey, Salles, and Weymark 2008, pp. 263−80.

Fleurbaey, M., M. Salles, and J. A. Weymark (eds) 2008: *Justice, Political Liberalism and Utilitarianism*. Cambridge: Cambridge University Press.

Forster, Malcolm R. 2001: 'The New Science of Simplicity'. In Zellner, Keuzenkamp, and McAleer 2001, pp. 83−119.

Forster, Malcolm R. and Elliott Sober 1994: 'How to Tell When Simpler, More Unified or Less Ad Hoc Theories Will Provide More Accurate Predictions'. *The British Journal for the Philosophy of Science*, 45, pp. 1−35.

Gaertner, Wulf 2006: *A Primer in Social Choice Theory*. Oxford: Oxford University Press.

Horwich, P. (ed.) 1993: *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MA: MIT Press.

Howson, Colin and Peter Urbach 1992: *Scientific Reasoning: the Bayesian Approach*. La Salle: Open Court.

Kelsey, David 1987: 'The Role of Information in Social Welfare Judgements'. *Oxford Economic Papers*, 39, pp. 301−17.

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

———1977a: ‘Objectivity, Value Judgment and Theory Choice’. In Kuhn 1977b, pp. 320−39.
———1977b: *The Essential Tension*, Chicago: University of Chicago Press.
Lipton, Peter 1990: *Inference to the Best Explanation*. London: Routledge.
McMullin, Ernan 1993: ‘Rationality and Paradigm Change in Science’. In Horwich 1993, pp. 55−78.
Newton-Smith, William H. 1981: *The Rationality of Science*. London: Routledge.
Roemer, John 1997: *Theories of Distributive Justice*. Oxford: Oxford University Press.
Sen, Amartya 1969: ‘Quasi-Transitivity, Rational Choice and Collective Decisions’. *Review of Economic Studies*, 36, pp. 381−94.
———1970: *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
———1977: ‘On Weights and Measures: Informational Constraints in Social Welfare Analysis’. *Econometrica*, 45, 7, pp. 1539−72.
———1986: ‘Social Choice Theory’. In Arrow and Instiligator 1986, pp. 1073−181.
Thagard, Paul 1978: ‘The Best Explanation: Criteria for Theory Choice’. *Journal of Philosophy*, 75, pp. 76−92.
Tsui, Kai-yuen and John A. Weymark 1997: ‘Social Welfare Orderings for Ratio Scale Measurable Utilities’. *Economic Theory*, 10, 2, pp. 241−56.
Zellner, A., H. Keuzenkamp, and M. McAleer (eds) 2001: *Simplicity, Inference and Modelling*. Cambridge: Cambridge University Press.