

**S. L. HURLEY**

**SUPERVENIENCE AND THE POSSIBILITY OF COHERENCE\***

(on intrapersonal analogues of Arrow's Theorem)

Mind, 1985, pp. 501-525. See my Natural Reasons (Oxford University Press 1989) for a revised version.

**1. Coherence accounts and the existence of coherence functions**

Consider views that involve claims of the following kind: to say that a certain act ought to be done is to say that it is favoured by the theory, whichever it may be, that gives the best account of the relationships among the various specific reasons (such as moral values, or legal doctrines and precedents) that apply to the alternatives in question. I will call any view of this kind a *coherence account*. Such views are held by many and are regarded by many others as a live option.<sup>1</sup> I will not here defend the assumption that views of this kind are attractive; rather I shall be concerned with whether or not a certain kind of challenge to such views is successful. Since few will have been worried by such challenges to begin with, the claim that they do not in the end succeed will come to few as a great relief. What is interesting, though, is how close they come to succeeding and what is needed to stop them,

---

\* I am grateful to many people for helpful criticisms of various drafts and comments in conversation, including Simon Blackburn, John Broome, Ronald Dworkin, William Ewald, Allan Gibbard, John McDowell, Derek Parfit, Christopher Peacocke, Kevin Roberts, Paul Seabright, Amartya Sen, and John Vickers; needless to say, all remaining errors are my own.

1. See and compare Norman Daniels, 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy*, 1979, pp. 256—82; Ronald Dworkin, 'No Right Answer?', in P. M. S. Hacker and J. Raz, eds., *Law, Morality and Society*, Oxford, Clarendon Press, 1977, pp. 58—84; 'Seven Critics', II *Georgia Law Review* 1205 (1977), p. 1258 *Taking Rights Seriously*, London, Duckworth, 1977 pp. 87, 104-9 119-22 126-7 159-68 283; Joel Feinberg, 'Justice, Fairness and Rationality', 81 *Yale Law Journal* 1004 (1972), section III Neil MacCormick, *Legal Reasoning and Legal Theory*, Oxford, Clarendon Press, 1978 chapter VII; John Rawls, *A Theory of Justice*, Cambridge, Mass., The Belknap Press, 1971, pp. 20ff., 48 ff.; Israel Scheffler, 'On Justification and Commitment', *Journal of Philosophy*, 1954, pp. 180-90.

namely, a requirement of supervenience. Many virtues have been claimed for supervenience doctrines; here is another.

According to coherence accounts, the deliberator's task is to seek moral, or legal, theories that display coherence. Another way of putting this is to say that deliberation involves a process of constructing hypotheses about the content of a *coherence function*, which takes us from rankings of alternatives under specific reasons to all-things-considered rankings in a way that meets certain conditions. Examples could be offered to support the plausibility of such an account. But it is one thing to claim the plausibility of such an account of what deliberators try to do, and another to claim that the coherence functions they seek exist. In order to support the latter claim we must say more about the conditions that coherence functions must meet. Until we have done so we cannot dismiss the possibility that, even if our account of their search is correct, deliberators are seeking a chimera.

If a coherence account is to permit us to defend the truth of any claim about what ought to be done, the minimal conditions reasonable to impose on a coherence function must be consistent. According to a coherence account, to claim that a certain act ought to be done is not to say that it is favoured by any given theory, or to say which theory does the best job of displaying coherence, but rather to say that the act is favoured by *the best theory that displays coherence*; the theory thus described is claimed to exist, but which theory it is must be discovered *a posteriori*. The claim is false if the theory that in fact is the best theory displaying coherence does not favour the act in question. But the claim is also false, *inter alia*, if there is no theory that displays coherence because coherence is impossible to obtain. It may be true that deliberators seek coherence but that all of their claims about what ought to be done are false because there is no such thing as coherence. I shall argue that what at first looks an alarmingly strong case for the non-existence of coherence functions does not in the end succeed.

## **2. The analogy between deliberation and social choice**

It is often assumed that the possibility of rationally resolving one – person conflicts is less problematic in principle than the possibility of rationally resolving conflicts between

persons. Bernard Williams has expressed scepticism about this assumption. He points out that

some one-person conflicts of values are expressions of a complex inheritance of values, from different social sources, and what we experience in ourselves as a conflict is something which could have been, and perhaps was, expressed as a conflict between two societies. . . The same point comes out the opposite way round, so to speak: a characteristic dispute about values in society, such as some issue of equality against freedom, is not one most typically enacted by a body of single-minded egalitarians confronting a body of equally single-minded libertarians, but is rather a conflict which one person, equipped with a more generous range of human values, could find enacted in himself.<sup>2</sup>

A person's various values may conflict with respect to alternatives with the same non-evaluative characteristics. The analogy between one – person conflict and many-person conflict may be taken to support efforts to model techniques for resolving many-person conflicts on techniques for resolving one-person conflict, on the assumption that the former techniques are no more problematic than the latter.<sup>3</sup> But the analogy may also be turned on its head: it may be taken to support scepticism about the possibility of rationally resolving one-person conflicts, on the assumption that this possibility is no less problematic than the possibility of rationally resolving many-person conflicts. Indeed, in the wake of Kenneth Arrow's impossibility result for social welfare functions, Kenneth May described an analogous impossibility result for the individual deliberator.

Arrow was concerned with the view that social welfare is some function of the conflicting preferences of individuals. He considered the problem faced by a society that tries to aggregate rankings of alternatives by individuals to get a ranking that reflects social welfare, and demonstrated that '...we cannot count on transitivity of group preferences even if individual preferences are transitive'. May reinterpreted Arrow's result to apply to the problem faced by an individual who must arrive at all-things-considered evaluations of alternatives that he ranks in different ways according to different criteria, and concluded

---

2. Bernard Williams, 'Conflicts of Values', pp. 222–3, in Alan Ryan, ed., *The Idea of Freedom*, Oxford, Oxford University Press, 1979, at pp. 221–32.

3. For example, see R.M. Hare in *Moral Thinking*, Oxford, Clarendon Press, 1981, pp. 109-10, and J. C. Harsanyi in 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', at p. 280 in Edmund S. Phelps, ed., *Economic Justice*, Harmondsworth, Penguin, 1973.

that ‘...we cannot expect individual preferences to be always transitive’ even given the transitivity of the rankings of alternatives effected by each criterion.<sup>4</sup> That is, May was concerned with the possibility of functions formally analogous to social welfare functions, which take as arguments rankings of alternatives generated by various conflicting criteria rather than by various individuals. He claimed that no function exists that meets conditions analogous to Arrow’s and takes us from transitive criterial rankings of alternatives to a transitive all-things-considered ranking of them.

Consider the following case. Suppose that the deliberator is a doctor who can treat only one of three urgent cases of the same illness; his alternatives are to treat *a*, *b*, or *c*. The illness has very similar symptoms and effects in the three cases, and the treatment would be equally successful applied to any of them, *a*, *b*, and *c* are all free of family obligations, *a* has had bad health all his life, and has several other serious conditions; *b* has another, somewhat less serious complaint, while the illness in question is *c*’s only complaint. *b* is the doctor’s own patient of long standing, *c* has been referred to him by a colleague, and *a* is a foreigner attending a conference of mathematicians in the doctor’s vicinity (which *b* and *c* are also attending). Finally, *c* promises to make brilliant advances in some branch of pure mathematics if cured of the disease, while *a* has somewhat less talent and *b* somewhat less still. The deliberator determines that considerations of equality, entitlement, and excellence or perfectionism have differential bearing on the alternatives, and rank the alternatives as follows:

equality	entitlement	excellence
a	b	c
b	c	a
c	a	b

Thus, the relevant criteria in this example are equality, entitlement, and excellence, and the alternatives are ranked with respect to each criterion according to the degree to which they satisfy it; the criteria are directional, such that any one alternative’s ranking above a

---

4. Kenneth O. May, ‘Intransitivity, Utility, and the Aggregation of Preference Patterns’, *Econometrica*, 1954, pp. 1-13 at p. ; see also pp. 3, 5-7,9-10. To. See also and compare Amos Tversky, ‘Intransitivity of Preferences’, *Psychological Review*, 1969, pp. 311-48; R. Duncan Luce and Howard Raiffa, *Games and Decisions*, New York, Wiley, 1966, pp. 318, 342 if., and David Kelsey, *Topics in Social Choice*, Oxford D.Phil. thesis. 1082.

second with respect to a criterion counts *per se* in favour of the first (see condition P\* below). The pattern of conflict the deliberator faces here is the same as that found in *voters' paradoxes*: If the rankings of alternatives were effected by groups of voters of equal number, instead of by considerations of equality, entitlement, and excellence, then the method of majority decision would produce an intransitivity. Alternative *a* would be ranked higher than alternative *b* by two out of three voters, alternative *b* would be ranked higher than alternative *c* by two out of three voters, and alternative *c* would be ranked higher than alternative *a* by two out of three voters. Thus, rankings of alternatives arrived at by the method of majority decision may be intransitive, even given the transitivity of rankings of alternatives by individual voters.<sup>5</sup>

Arrow proved that there is no function from transitive individual orderings of alternatives to a transitive social ordering that meets the following conditions:

*P (Weak Pareto Principle)*: For any alternatives *x* and *y*, if all individuals prefer *x* to *y*, then society prefers *x* to *y*.

*D (Non-Dictatorship)*: There is no individual such that, for all alternatives *x* and *y*, if that individual prefers *x* to *y* then society prefers *x* to *y*.

*I (Independence of Irrelevant Alternatives)*: For any alternatives *x* and *y*, if the preferences of all individuals as between *x* and *y* remain the same then the preference of society as between *x* and *y* remains the same; that is, the preferences of society as between *x* and *y* depend only on the preferences of individuals as between *x* and *y*.

*U (Unrestricted Domain)*: For any set of alternatives and any set of individuals, the domain of the social welfare function includes all orderings of alternatives by individuals.

There is no method of aggregating transitive individual preferences that meets these conditions and that guarantees transitive social preferences; Arrow considered this a generalization of the possibility of 'collective irrationality' revealed by the voters' paradox.<sup>6</sup>

---

5. Larger voters' paradoxes may be constructed in which the ratio of voters favouring one of a pair of alternatives to those disfavouring it approaches unity.

6. See Kenneth J. Arrow, *Social Choice and Individual Values*, 2nd edition, New Haven and London, Yale University Press, 1963 chapter 3, pp. 59–60, 96–7 *if.*; compare Amartya K. Sen *Collective Choice and Social Welfare*, San Francisco, Holden Day, 1979, pp. 37-8; *Choice, Welfare and Measurement*, Oxford. Basil Blackwell, 1982, pp. 164-5, 230-1.

His proof works by showing that a dictator will emerge from any method of aggregation that produces a transitive social ordering and that meets conditions P, I, and U. Arrow interprets his result to mean that ‘. . . the doctrine of voters’ sovereignty is incompatible with that of collective rationality.’<sup>7</sup> The appropriateness of the conditions Arrow imposes on a social welfare function, and thus the significance of his result for democratic theory, are controversial, especially in light of the conditions’ effect of ruling out many methods of arriving at interpersonal comparisons of utility.<sup>8</sup>

However, I shall not be considering the appropriateness of Arrow’s conditions or the significance of his result as originally interpreted, but rather the appropriateness of placing conditions analogous to Arrow’s on deliberation and the significance of May’s reinterpretation of Arrow’s result for coherence accounts. Thus, I revert to the original formulation of the doctor’s problem as a problem of deliberation. The doctor may consider a method of deliberation analogous to the method of majority decision, which would recommend the alternative that is superior according to a majority of criteria. However, this method recommends  $a$  over  $b$ ,  $b$  over  $c$ , and  $c$  over  $a$ . May suggests that this possibility of deliberative irrationality may be generalized: intransitive all-things – considered recommendations may be generated by any method of deliberation that meets conditions analogous to Arrow’s, with criteria replacing individuals in the formulation of the conditions.<sup>9</sup>

We may regard the majority-of-criteria method as crude and unappealing, and so not be alarmed at its inadequacy. But the claim that there is no function from orderings of

---

7. Arrow, op. cit., p. 60.

8. See Sen, *Collective Choice and Social Welfare*, chapter 7.

9. May works with variations of Arrow’s original conditions, rather than with the now familiar versions P, D, I, and U, whose analogues I consider. However, the essential point of May’s article is to suggest that Arrow’s result may be interpreted to apply to conflicting criteria as well as to individuals with conflicting preferences, rather than a technical point about the proof of the result in its new application. The semantic alteration does not affect syntactic relationships among the conditions in either the original set or the now familiar set. Moreover, the crucial critiques of independence and unrestricted domain apply equally to May’s versions of the conditions.

It may be that May never intended his suggestion to be applied to criteria that have the properties of values, in particular, supervenience. In so applying it I am not attempting to interpret or criticize May, but only to follow out a line of thought that his suggestion brings to light, whether he intended it to or not. The line of thought is interesting in its own right, since if the application can be made a familiar kind of appeal to coherence in moral philosophy may be undermined.

alternatives by specific criteria to an all – things-considered ordering that meets certain conditions may indeed be alarming. According to coherence accounts, what ought to be done is some function of the information provided by the various specific reasons that apply to the alternatives in question; deliberation can be represented as a process of constructing hypotheses about the content of the function. If May's reinterpretation of Arrow's framework accurately represents the situation of a deliberator who pursues coherence in the face of conflicting reasons, then there is no question but that the coherence functions deliberators seek do not exist. Conversely, if coherence functions do exist, then at least one of the conditions analogous to the conditions that produce Arrow's result does not represent a constraint faced by a deliberator in pursuit of coherence; in order to vindicate a coherence account, we must show that one of the conditions fails. Which one, however, is not obvious. Thus we must scrutinize the conditions that give rise to May's claim.

Since Arrow proved his theorem, many other impossibility results have been obtained. I will consider the analogues of three sets of conditions that produce impossibility results for social welfare functions; these results seem, for reasons that will emerge, especially threatening to coherence accounts. (Of course, I do not claim that there is no threat from any other impossibility result; there may well be threats I have not noticed.) For each set of analogous conditions, if the conditions in it are conditions of rational deliberation, then non-dictatorial coherence functions do not exist and we must conclude that, despite their attractions, pluralistic theories (or, more precisely, non-lexicographic forms of pluralism) cannot be coherent. The scope of such a conclusion would be disturbing: for example, not even Rawls's theory is thoroughly lexicographic, as the priority of liberty only sets in after society reaches a certain stage of development.<sup>10</sup>

### **3. The analogues of conditions P and D**

The first set of conditions to be scrutinized, then, is the set of conditions analogous to Arrow's but with criteria playing the role of individuals. The rankings of alternatives by criteria are required to be orderings, hence transitive. The conditions in this set are:

---

10. See Rawls, *op. cit.*, pp. 151–2, 542.

*P\** (*Dominance*): For all alternatives x and y, if all criteria rank x above y then x ranks above y all things considered.

*D\** (*Non-Dictatorship*): A coherence function must not give so much weight to one criterion that it outweighs any criterion that conflicts with it under any circumstances; that is, it must not be the case that there is one criterion such that any one alternative's superiority over any other according to this criterion always results in its superiority all things considered, regardless of how other criteria rank those alternatives.

*J\** (*Independence of Irrelevant Alternatives*): For all alternatives x and y, if the rankings of x and y by all criteria remain the same then the ranking of x and y all things considered remains the same; that is, a deliberator's ranking of a pair of alternatives all things considered depends only on the ranking of those alternatives by all relevant criteria, and not on the rankings of other alternatives.

*U\** (*Unrestricted Domain*): For any set of alternatives and any set of criteria, the domain of the coherence function includes all orderings of alternatives by criteria.

*P\** just amounts to a principle of dominance, which is not controversial. Against the weak Pareto principle itself well-known objections have been raised by Amartya Sen, but these run to the privileged role the principle gives to one kind of information, welfare information.<sup>11</sup> The analogue of the weak Pareto principle, *P\** involves no presumption in favour of reasons of some particular kind. It merely says that if all the applicable reasons, whatever they may be, favour one alternative over another, then their unanimous recommendation must be followed, and the non – evaluative characteristics of the alternatives cannot affect their ranking all things considered. This is an unobjectionable condition to place on coherence functions.

What about the analogue of the non-dictatorship condition, *D\**? May suggests that if one of our conflicting criteria were to dictate, there would be no deliberative problem.<sup>12</sup> If one criterion were to outweigh all others in all circumstances, the sought-after function would be obvious; all our resolutions of past and hypothetical conflict cases, which the theory is

---

11. See Sen, *Collective Choice and Social Welfare*, chapters 6 and 6\*; *Choice, Welfare and Measurement*, pp. 248ff., 285–326, 341–6.

12. See May, *op. cit.*, p. 12.

supposed to account for in terms of the relationships among the relevant criteria, would reflect its domination. In assuming that there is a deliberative problem we are assuming that no one of the criteria we start with does dictate.

But this is too swift. Someone might object as follows. The best theory we are able to construct may tell us that some of our resolutions of conflicts have been mistaken. Why should a theory not tell us that all resolutions of conflicts that do not reflect the dictatorship of a particular criterion have been mistaken?

A theory could tell us this, but familiar theoretical norms suggest that it could not be the best possible theory if it did. By telling us that all such resolutions have been mistakes, a theory that establishes the dictatorship of a particular criterion achieves simplicity, admittedly a theoretical virtue, but at too great a price in terms of the data the theory is supposed to account for. We will always do better to look for a less revisionary theory, one that does not regard so much of the data as flawed as a theory establishing dictatorship does. Condition  $D^*$  reflects this constraint on theorizing.

Alternatively, we may impose  $D^*$  simply because we are interested in whether non-dictatorial theories that meet our other conditions are possible; if they are not, it is worth knowing.

Note that  $D^*$  does not rule out utilitarian or other monistic theories. Sophisticated forms of utilitarianism appeal to rules, motives, and/or a separation of levels in a way that gives utility a role to play within a theory, rather than merely asserting its domination in all conflicts. Such forms of utilitarianism do not simply overrule the claims of justice, for example, but try to account for them within a structured utilitarian theory – at least to a degree that theoretically justifies overriding residual claims that cannot be so accommodated. In this sense, a theory that is monistic may nevertheless recognize a plurality of values. On a coherence account, utilitarianism is subject, irrespective of its monistic character, to the same standards *qua* theory as non-monistic theories; if a form of utilitarianism is to succeed, it must dominate other theories on theoretical grounds. Nothing in  $D^*$  rules out the possibility that some form of utilitarianism may do so.  $D^*$  merely reflects the reasonable view that any theory which simply asserted the domination of some

particular value, unqualified in respect of conflicting values, would be so weak as a theory as to be out of the running; but utilitarianism need not be such a theory.

A theory's success in accounting for settled cases in which values conflict provides theoretical justification for applying it to resolve unsettled cases in which those values conflict. But this justification should not be confused with the dictatorship of one of the values the theory is about. If this confusion is made,  $D^*$  alone rules out the possibility of a coherence function, as any successful function could immediately be reinterpreted as a dictatorial argument of a simpler function and thus debarred. The analogous confusion with respect to social welfare functions does not arise because functions of individuals' preferences have no tendency to materialize into individuals whose dictatorial status would be a matter of concern. We may admit that there is considerable interplay between various specific reasons and our theorizing about them, so that the distinction between practical and theoretical reasons is not absolute and there are differences of degree; but this interplay does not render values so obligingly nascent as to obliterate the distinction entirely. A particular hypothesis about the relationships among practical reasons may come to have the discrete character and force of a practical reason itself, but this result is not an instant consequence of theorizing, and may require conceptual, perceptual, and phenomenological development.

#### **4. The analogue of condition I and supervenience**

Condition  $J^*$  has two aspects,<sup>13</sup> only one of which is essential to an impossibility result.

The *ordinalist aspect* restricts the quality of information provided by a criterion about any pair of alternatives to ordinal information. Of course, we may well wish to make use of cardinal information about the degree to which the alternatives satisfy a particular criterion and hence object to the ordinalist aspect of  $J^*$ . However,  $J^*$  can be reformulated to admit cardinal information without affecting the impossibility result, since such cardinality with respect to different criteria does not *per se* provide intercriteria comparisons of alternatives, any more than cardinal utility functions *per se* provide interpersonal comparisons.<sup>14</sup>

---

13. See Sen, *Collective Choice and Social Welfare*, pp. 89–90.

14. In the social choice literature, condition I is reformulated to allow for cardinal information within the framework of *social welfare functionals*, which may take as arguments cardinal non-comparable

The *irrelevant alternative aspect* proper is crucial in precipitating the impossibility result, whether in conjunction with ordinal or cardinal information. It simply says that the ranking of two alternatives all things considered is determined solely by information of the admissible form provided by the applicable criteria about those alternatives, and no others; it restricts the quantity, but not the quality of admissible information. This restriction seems difficult to resist: if all admissible criterial information that does concern a pair of alternatives has been taken into account, how could it be helpful to consider criterial information about other alternatives?<sup>15</sup>

But perhaps it can be argued that a coherence account does require us to resist this restriction: in trying to determine what consistent set of relationships may obtain among our various values – that is, what coherence amounts to – we characteristically appeal to settled and hypothetical cases other than those at issue; this is what we must do if we are to give any particular content to the idea of coherence. We analyse such cases in order to determine the relationships among the relevant values in various circumstances. If we were to find the irrelevant alternatives ranked otherwise by the relevant values than they are in fact, our theory of the relationships among those values and the weights we assign to them under various circumstances might differ as a result, so that our all – things – considered evaluations of the alternatives at issue might vary with the rankings of irrelevant alternatives. Thus, J\* is violated by a characteristic feature of deliberation conceived as a search for coherence; the possibility of coherence is vindicated by the failure of 1\* to represent a reasonable constraint on deliberation.

There are at least two problems with this argument against 1\* as it stands. In order for 1\* to be violated, there must be a pair of alternatives a and b such that the rankings of a and b

representations of individuals' preferences. See Sen's conditions R, C and M, *Collective Choice and Social Welfare*, chapter 8\*.2 see also chapter 8.2, and his *Choice, Welfare and Measurement*, pp. 230-1. For simplicity I have stated the conditions and their analogues in terms of criterial orderings of alternatives, but my arguments do not preclude richer criterial information.

It is not an objection to condition 1\* reformulated to allow coherence functionals to take as arguments cardinal representations of criterial information that such representations do not provide intercriteria comparisons. It is the job of the coherence *function*, not its *arguments*, to provide an account of the relationships among conflicting values and the circumstances under which certain values outweigh others. The question is whether or not 1\* represents a reasonable constraint on functions whose job is to provide such comparisons.

15. See and compare May, *op. cit.*, pp. 9–10; Arrow, *op. cit.*, pp. 27–8; Sen, *Collective Choice and Social Welfare*, pp. 37–8.

by all criteria remain constant while nevertheless the all-things-considered ranking of a and b to which a given coherence function takes us varies, presumably with the criterial rankings of other, 'irrelevant' alternatives.

The first problem for the argument against 1\* is that, even if we were to grant the existence of a pair of alternatives a and b whose criterial rankings remain constant while their all-things – considered ranking varies with the criteria! rankings of other alternatives, such variation would reflect different hypotheses about the content of a coherence function, hence different functions. But for our purposes 1\* must be taken to constrain applications of any given coherence function, not the process of arriving at that function to begin with, to constrain relations between input and output of one function, not relations among different functions. Deliberation involves the search for a theory about our values as they actually are, not a methodology, a theory about theories. It proceeds by reference to hypothetical cases that may be counterfactual, but may not be so in a way that is counterevaluative; the function of the hypothetical cases is to reveal more about the values we actually have, not to introduce information about other possible values. In the context of moral theory, for example, a coherence function represents a substantive theory that takes the relations between given values as its subject matter. If our values had been other than what they are or had had different content, we might well need a different theory to account for the relationships between them.

1\* is indeed objectionable in application to a methodology as opposed to a substantive theory, as it would require different substantive theories, theories about different sets of different logically possible values, to rank two alternatives in the same way if the corresponding values in each set rank them the same way. This is implausible for reasons similar to those that would make imposing such a requirement on two different persons, with different sets of values, implausible. There are difficulties about what the correspondence of values across sets of different logically possible values could consist of. Moreover, such a requirement would mean that a value in one theory (or for one person) could not have more weight than its corresponding value in another theory (or for another person) as a result of its relationships to other values and its role in the theory (or the

person's conception of himself) as a whole, as revealed by consideration of alternatives other than the two at issue to which the values also apply. Theories about different values are in part differentiated by just such differences of structure and weight, which are in turn called for by differences in the values the theories are about. The contemplated restriction would indeed frustrate the basic purpose of theorizing about values by reference to hypothetical cases, namely, to arrive at the most coherent possible account of the relationships among those values in particular.

Suppose, then, that 1\* is not put forward as a condition on methodologies but a condition on substantive theories. Then the second problem for the argument against 1\* that I gave four paragraphs back is that, if the criteria involved are given values, the supervenience of evaluative judgements on non-evaluative judgements prevents the criteria! rankings of any given set of alternatives from varying at all. *Supervenience* says that, as a matter of logical necessity, if two alternatives differ in some evaluative characteristic, then they differ in some non-evaluative characteristic as well. 'The doctrine of the supervenience of values is well established, and I will not argue for it or defend it here. The ranking of a given set of alternatives by a given value may not be determined as a matter of logic; that is, the evaluative characteristic may not be logically reducible to non - evaluative characteristics, so that the counterfactual supposition that the value might have had some other content is not self-contradictory (though such counterevaluative suppositions are irrelevant to moral theorizing). But what is determined as a matter of logic, is that evaluative characteristics and rankings cannot vary independently of non - evaluative characteristics of alternatives, in the way that colour can vary independently of shape. Of course, two different values may conflict about alternatives with given non-evaluative characteristics; if they could not there would be no deliberative problem. But a given evaluative criterion cannot 'change its mind' about given alternatives; if it seems to, either the criterion has changed (i.e. there are two different criteria involved), or the non - criterial description of the alternatives, hence the alternatives themselves, have changed. Thus, the rankings by given values of any given pair of 'irrelevant' alternatives cannot vary. And, *ex hypothesi* when condition 1\* is at issue, the evaluative rankings of the alternatives at issue do not vary.

But if the criteria! rankings of all alternatives must remain constant, then the all-things-considered ranking of the alternatives at issue cannot vary either, as there is nothing for it to vary with. Of course we can consider different or larger sets of evaluative criteria, but I\* applies to a given pair of alternatives and a given set of criteria.<sup>16</sup>

Recall that the impossibility result is preserved if J\* is reformulated to allow for cardinal information with respect to each criterion. But super – venience guarantees 1\* in this guise as well, for it holds cardinal evaluative information about a given set of alternatives constant no less than ordinal evaluative information about a given set of alternatives. If cardinal evaluative information changes, then either ‘the evaluative criterion has changed or the non

---

16. Formally, supervenience may be stated as follows, where x and y range over alternatives, or possible acts, and ranges over non-evaluative specifications of an alternative’s characteristics: For any evaluative specification (including ordinal relational specifications and cardinal specifications) of an alternative’s characteristics  $\psi$  is logically necessary that  $\forall x \forall y ((\phi x \ \& \ -\phi y) \rightarrow \exists \psi (\psi x \ \& \ -\psi y))$ . I of course presuppose that different values may supervene on the same set of non—evaluative characteristics, and that such different values may conflict; thus different evaluative criteria may of course differ with respect to alternatives that have the same non-evaluative characteristics. Note that I do not claim that all criteria, in all decision problems, must be evaluative and hence logically supervenient; for example, I do not rule out the possibility that a decision problem might be structured so as to have a colour as a criterion, which distinguishes between two boxes identical except with respect to colour. (If colour supervenes it does so as a matter of physics, not logical necessity.) A criterion merely provides information, which may be ordinal or cardinal, about the degree to which alternatives have some criterial characteristic that counts for or against them; whether a criterial characteristic must supervene on other characteristics of the alternatives is a further question. However, evaluative criteria, for example justice, must be supervenient on non-evaluative characteristics. Our interest in decision problems that do involve evaluative criteria, for example in the context of Rawlsian reflective equilibrium, hardly needs defence. Furthermore, while values are supervenient criteria, supervenient criteria may not be values; there may be substantive as well as formal conditions on values, i.e. conditions on their content. See David Lewis, ‘New Work for a Theory of Universals’, *Australasian Journal of Philosophy*, 1983, pp. 375-7.

The supervenience of one set of concepts or properties on another is a much-discussed topic in moral philosophy and philosophy of mind. For a sample of the large literature see R. M. Hare, *Freedom and Reason*, Oxford, Clarendon Press, 1963, chapters 2, 3, and *Moral Thinking*, Oxford, Clarendon Press, 1981, chapters 5, 6, 7, 10 J. L. Mackie, *Ethics*, Harmondsworth, Penguin, 1977 chapter 4; Simon Blackburn, ‘Supervenience Revisited’, in Ian Hacking, ed., *Exercises in Analysis*, Cambridge, Cambridge University Press, 1985 and *Spreading the Word*, Oxford, Clarendon Press, 1984, chapter 6, and ‘Moral Realism’ in John Casey, ed., *Morality and Moral Reasoning*, London, Methuen, 1971, pp. 101-24 Jaegwon Kim, ‘Supervenience and Nomological Incommensurables’, *American Philosophical Quarterly*, 1978; John Haugeland, ‘Weak Supervenience’, *American Philosophical Quarterly*, 1982 J. Dancy, ‘On Moral Properties’, *Mind*, 1981; Harry A. Lewis, ‘Is the Mental Supervenient on the Physical?’, in Bruce Vermazen and Merrill B. Hintikka, eds., *Essays on Davidson*, Oxford, Clarendon Press, 1985 See Hare, Mackie, and Blackburn on the logical necessity of the supervenience of evaluative concepts. Note that supervenience does not entail the reduction of evaluative to non-evaluative characteristics (compare the supervenience of the characteristic of being a table on fundamental physical characteristics); see Blackburn and Kim.

In arguing that supervenience guarantees 1\*, I assume that alternatives with different non-criterial characteristics are different alternatives. Other assumptions could be made that would themselves guarantee J\* and that might seem to make the appeal to supervenience superfluous; but it is not superfluous when we come to reject condition R\*.

– evaluative characteristics of the alternatives have changed. Supervenience makes it logically impossible to violate J\*•

However, the sense in which 1\* is met seems to be somehow trivial; it is met because, given supervenience, it is empty. An analogous charge of triviality is found in the social choice literature dealing with impossibility results for Bergson – Samuelson social welfare functions. Bergson – Samuelson social welfare functions take as arguments any one set of preference orderings of individuals, but once the set is fixed, the individuals are not permitted to change their minds and other possible preference orderings are not considered. The approach to social welfare that holds individual preferences constant in this way while not imposing any restrictions on their content is known as the *single-profile approach*, in contrast to Arrow’s *multiple-profile approach*, which does allow individual preferences to vary.

The single – profile approach imposes a restriction on individual preferences analogous to that imposed on criterial rankings by supervenience (given my earlier point that coherence functions represent theories about values, not methodologies). As Samuelson explains, ‘. . . here we are in the domain of given individuals’ tastes and values and not in the wider and different Arrow domain of all that their tastes and values might be.’ And ‘...one and only one of the.. . possible patterns of individual orderings is needed. It could be *any* one, but it is only one.’ The Bergson – Samuelson social welfare function is not concerned with counterfactual possibilities about the content of preferences, but only with the content of actual preferences, whatever that may be, just as a moral theory is not concerned with what the content of the relevant values might have been, but only with what it is. Thus, perhaps the single – profile approach rather than the multiple – profile approach provides the correct analogy to deliberation involving values. It should not be surprising that Samuelson claims that condition I is automatically met by Bergson – Samuelson social welfare functions, that it is built in from the beginning.<sup>17</sup>

---

17. See Paul A. Samuelson, ‘Reaffirming the Existence of “Reasonable” Bergson-Samuelson Social Welfare Functions’, *Economica*, 1977, pp. 81–8 at p. 8z; and ‘Arrow’s Mathematical Politics’, in Sidney Hook, ed., *Human Values and Economic Policy*. New York, New York University Press, 1967 pp. 41-51 at pp. 43, 47, 48–9. See also Kevin W. S. Roberts, ‘Social Choice Theory: the Single-profile and Multi-profile Approaches’, *Review of Economic Studies*, 1980, pp. 441–50, at p. 443.

The charge of triviality is made by Sen, against the sense in which condition I is met by Bergson – Samuelson social welfare functions. Sen suggests that condition I is essentially an inter – profile constraint and must be reformulated in a way that is appropriate for the single-profile approach if we are properly to address the question whether impossibilities of the Arrow type arise for Bergson – Samuelson social welfare functions.<sup>18</sup> If Sen is right, an analogous reformulation of condition I\* is needed if we are properly to address the question whether coherence functions exist that take as arguments rankings of alternatives generated by conflicting values, which are constrained by supervenience.

The social choice literature contains two types of suggestion as to how the needed reformulation may be achieved, the first made by Robert Parks, and by Murray Kemp and Yew-Kwang Ng, and the second made by Kevin Roberts. These two types of suggestion contribute to the second and third sets of conditions leading to impossibility results whose analogues I shall be concerned with. The question becomes: do the analogues of conditions needed for single – profile impossibility results represent reasonable constraints on deliberation?

##### **5. The analogue of single-profile neutrality**

Parks and Kemp and Ng have obtained single-profile impossibility results that depend on a strong reformulation of condition I, which amounts to the outright assumption of neutrality with respect to non-welfare information:

*N (Neutrality with respect to Non- Welfare Information):* For any alternatives w, x, y, and z, if the preferences of all individuals are the same as between w and z as they are as between x and y, then the preference of society as between w and z must be the same as its preference as between x and y.<sup>19</sup>

---

18. See Sen, *Choice, Welfare and Measurement*, pp. 251–6.

19. See Murray C. Kemp and Yew-Kwang Ng, ‘On the Existence of Social Welfare Functions, Social Orderings and Social Decision Functions’, *Economica*, 1976 pp. 59–66, at p. 60; and Robert P. Parks, ‘An Impossibility Theorem for Fixed Preferences: A Dictatorial Bergson-Samuelson Welfare Function’, *Review of Economic Studies*, 1976 pp. 447-50, at p. 448. See also Kemp and Ng, ‘More on Social Welfare Functions: The Incompatibility of Individualism and Ordinalism’ *Economica*, 1977, pp. 89-90.

A condition of neutrality such as N operates to deprive us of information. In the social choice context, neutrality with respect to non-welfare information deprives us of all but welfare information about the alternatives. In the deliberative context, neutrality with respect to non-criterial information deprives us of all but criterial information about the alternatives. The analogue of N is:

*N\** (Neutrality with respect to Non-Criterial Information): For any alternatives w, x, y, and z, if all criteria rank w and z in the same way they rank x and y, then w and z must be ranked in the same way, all things considered, as x and y.

In order for a neutrality condition to be other than trivially satisfied it is not necessary for the ranking of given alternatives by any individual or any criterion to vary, as variation may be provided by the characteristics of the alternatives; the characteristics of w and z may differ from the characteristics of x and y even though all individuals, or criteria, rank w and z in the same way as x and y. Neutrality with respect to all but information of the favoured kind requires that such pairs of alternatives be treated the same way in the final ranking. The difference in logical form that gives the neutrality conditions bite where the independence conditions had none is as follows (where ' $x \frac{1}{i} R y$ ' means that in situation I person or criterion i ranks x at least as high as y):

$$\forall x \forall y < \forall i \left( \left( xR \frac{1}{i} y \leftrightarrow xR \frac{2}{i} y \right) \& \left( yR \frac{1}{i} x \leftrightarrow yR \frac{2}{i} x \right) \right) \\ \text{Independence:} \quad \rightarrow (xR^1 y \leftrightarrow xR^2 y) >$$

$$\text{Neutrality:} \quad \forall w \forall x \forall y \forall z < \forall i \left( (xR_i y \leftrightarrow wR_i z) \& (yR_i x \leftrightarrow zR_i w) \right) \\ \rightarrow (xRy \leftrightarrow wRz) > .$$

Much of the labour in Arrow's proof of the emergence of a dictator is spent on deriving elements of neutrality from conditions U, P, and I, which are independently attractive as conditions on a social welfare function; if neutrality is simply assumed, it is not surprising that an impossibility result follows.<sup>20</sup> We might think that single-profile impossibility results

20. Sen makes this point in unpublished material. The close relationship between D\* and the failure of N\* in the context of legal deliberation shows up in the tension, familiar to lawyers, between the doctrine of *stare*

that help themselves to neutrality from the beginning are less disturbing than Arrow's result: it is only collectively that the individually attractive conditions U, P, and I impose the informational parsimony that condition N imposes explicitly, thereby making itself unattractive as a canonical constraint on a social welfare function. Analogous thoughts about N\* require us to look carefully at its merits as a condition on reasonable deliberation.

The immediate effect of N\* is to rule out the use of non-criterial information, parallel to the effect of N in ruling out the use of non-welfare information. So long as, for each relevant criterion, the information about one pair of alternatives with respect to that criterion is the same as the information about another with respect to that criterion, the resolution of criterial conflict must be the same for both. One criterion cannot outweigh the other in application to alternatives of a certain non-criterial character while the second outweighs the first in application to alternatives of a different non-criterial character. The theory must assign weights to the criteria that are independent of the non-criterial character of the alternatives.

Thus an objection to N\* might be that it requires theories about the relationships among values to be gratuitously crude. We might well want to distinguish different types of circumstance in describing these relationships without being forced to count such circumstances as having independent reason-giving force. I shall develop this point by means of three counter-objections that lead to refinements of the objection to N\*.

Consider the following appeal to non-criterial information. The question before a judge is whether or not to admit evidence of two of a defendant's prior felony convictions, both over 10 years old, for the purpose of aiding the jury in evaluating the credibility of the defendant's testimony. A reason to do so is that the evidence is relevant and hence generally admissible under the rules of evidence. A reason not to do so is that there is a danger of prejudice to the defendant if the jury comes to regard him as a hardened criminal and hence not to be concerned about the possibility of mistakenly convicting him of the crime he is

*decisis* and the requirement of 'reasoned distinction': greater resistance to overruling earlier decisions results in greater tolerance of rather more arbitrary distinctions between cases. See Henry M. Hart, Jr. and Albert M. Sacks, *The Legal Process*, Cambridge, Mass., Harvard Law School mimeograph, 1985, problems 12 and 21. I defend D\* on *stare decisis* grounds, in effect, i.e. resistance to regarding too many earlier resolutions as mistaken, and thus the possibility of non—neutrality is admitted.

being tried for; the rules of evidence require exclusion of evidence when its potential for prejudice outweighs its probative value. These two reasons are balanced against one another by a general rule that excludes evidence of prior convictions more than 10 years old on grounds that their potential for prejudice is likely to outweigh their probative value. However, the judge may reason as follows. By contrast to the cases for which the general rule was developed, the defendant in this case is being tried for the crime of being a convicted felon in possession of firearms. In these circumstances, ' . . . the prejudicial impact of proof of prior convictions is considerably lessened because the jury already knows the defendant had a record. At the same time the value of the evidence to the jury in determining the credibility of the defendant as a witness is somewhat enhanced because a man with a more extensive record is much more likely to know it is unlawful to possess weapons and to guard against the danger.'<sup>21</sup> The information about the distinctive circumstances of the case that the judge appeals to in adjusting the relative weights of the reasons in play does not count *per se* as a reason for or against admitting evidence, but rather affects the relationship between the reasons. Thus the objection to N\* is that to rule out the appeal to such information would be to reduce drastically and arbitrarily the possibility of finesse in our account of such relationships.

However, this objection is open to a counter-objection: appeals to apparently non-reason-giving circumstances may and should be interpreted as qualifying and refining the reasons we recognize in the first place rather than either as invoking independent reasons or as contributing to a theory about the relationships among unqualified reasons. For example, Sidgwick writes:

...it appears that a clear *consensus* can only be claimed for the principle that a promise, express or tacit, is binding, if a number of conditions are fulfilled: viz, if the promisor has a clear belief as to the sense in which it was understood by the promisee, and if the latter is still in a position to grant release from it, but unwilling to do so, if it was not obtained by force or fraud, if it does not conflict with definite prior obligations, if we do not believe that its fulfilment will be harmful to the promisee, or will inflict a disproportionate sacrifice on the promisor, and if circumstances have not materially changed since it was made. If any of these conditions

---

21. Engel, J., dissenting in *U.S. v. Sims*, 558 F.2d 1145 (1978).

fails, the *consensus* seems to become evanescent, and the common moral perceptions of thoughtful persons fall into obscurity and disagreement.<sup>22</sup>

It may be that certain of the circumstances Sidgwick mentions figure in essential qualifications of any obligation to keep promises; we appeal to such circumstances in the course of theorizing about the value of keeping promises and refining our view of it.

In responding to this counter-objection we sharpen the original objection to N\* as follows. We can admit that some of our theorizing in effect qualifies or revises our values without being forced to reconstrue all our theorizing about values as theorizing of this kind; the suggested defence of N\* fails to recognize the limits to such qualification and revision. Imagine a case in which the obligation to keep a promise qualified as Sidgwick suggests comes into conflict with the value of human life. When a promise is made the promisor correctly calculates that keeping it will create an extremely small risk to the life of a third party. After it is made, however, an unlikely event occurs, so that keeping the promise will create a substantial risk to the life of a third party. When we consider comparable circumstances in which the overriding of one value by the other recommends itself, and analyse the character of the promises involved and the character of the threat to life involved, we are not contributing to an elaborate qualification of the obligation to keep promises or the value of human life or both so that they no longer conflict at all.

It may well be formally possible to reconstrue a successful hypothesis about the content of a coherence function as a single all-purpose highly refined criterion, but this possibility is not to the point. When relevant criteria are specific moral values or legal doctrines or precedents they have discrete and elastic identities, and are not indefinitely malleable; such values provide substantive constraints on the contents of our theories which may not always yield to formal considerations. If the problem is how to resolve a conflict between such values, it is not solved by assuming we have got a resolution and inventing a criterion to sum it up. At some point theorizing about the relationships among such values ceases to qualify away conflict and takes it as its subject matter. In at least some cases even fully qualified values stand in conflict with one another, so that anything we do will infringe one

---

22. Henry Sidgwick, *The Methods of Ethics*, 5th edition, London, Macmillan, 1893, p. 311.

of them. In these cases we decide what to do not by distorting our values beyond recognition, or by ignoring them and adopting new ones, but by scrutinizing them, and other cases of conflict in which they apply, to discern some orderly and consistent set of relationships among them. For any such theory or coherence function, the value of deciding what to do in conflict cases according to that theory is just as distinct from the values the theory is a theory about as they are from one another.

But this sharpened objection to  $N^*$  is in turn open to a counter-objection, namely, that it misrepresents the considerations that make  $N^*$  appropriate. These considerations need not involve the view that all appeals to non-criterial information serve to qualify away conflict. We may appeal to non-criterial information in the course of determining the extent to which alternative acts or outcomes are supported or discouraged by the relevant criteria, the intra-criterial positions of the various alternatives; the criteria in question may conflict nevertheless. We may qualify and revise our criteria so that our use for non-criterial information is reflected in what the criteria themselves tell us in various circumstances rather than in the weight we give to what they tell us in various circumstances, without going all the way to a single, dictatorial if highly refined criterion. That non-criterial information has a role in determining what the criteria tell us about the alternatives does not imply that it has a further role in determination of the weight attached to the various criteria. Indeed, it might be claimed that any assignment of different weights to criteria under different circumstances could be reconstrued as a uniform weighting of criteria plus an adjustment in the information provided by the criteria about the different circumstances.

Again, there is justice in this objection, but it goes too far. Many appeals to non-criterial information are of the kind described by the objection. But it would rule out other uses of such information that there is no reason to rule out. Again we can respond by sharpening the objection to  $N^*$ . Even if the claim about reconstrual is correct as a formal matter, it is beside the point. Such technical reconstrual may not correctly describe the role of non-criterial information in theories about given criteria, in particular, about specific familiar values. That is, it will not always be true that the circumstances appealed to alter the degree to which the relevant values support or discourage the alternatives rather than the weight

we attach to the values; relationships among the specific familiar values we are interested in may be holistic, so that more or less weight is given to values such as excellence, autonomy, or equality depending on the extent to which they, and other values, are already satisfied.

Here is an example to which the refined objection to  $N^*$  applies. Suppose that we are trying to decide how resources ought to be distributed, and that conflicting considerations of total welfare, of equality of resources, of equality of welfare, and of responsibility for certain personal traits are admitted. As a result of deliberation we might arrive at a pluralistic theory that tells us to distribute resources so as to maximize welfare except in circumstances when by doing so we would leave someone with less than a certain minimum level of resources or would leave a handicapped person with less than a certain minimum level of welfare (a handicapped person is someone with a characteristic for which he is not responsible that causes him to be brought to a much lower-than-average level of welfare by any given level of resources); in these circumstances the theory tells us to distribute resources so as to maintain a certain roughly equal minimum level of resources for everyone and a certain roughly equal minimum level of welfare for the handicapped, respectively.

What is the role in such a theory of the information that a distribution would leave someone at a certain level of resources? It does not tell us that considerations of equality of resources favour it more or less; the distribution may involve more or less inequality of resources than some other, which leaves no one at that level. Rather, it tells us that circumstances obtain in which the theory assigns certain relative weights to considerations of welfare and considerations of equality of resources. Parallel comments apply to the role of the information that a distribution would leave a handicapped person at a certain level of welfare. This information does not tell us that considerations of equality of welfare favour the alternative more or less; again, it may involve more or less inequality of welfare than another, which leaves no one at that level. Nor does it tell us that considerations of responsibility favour the alternative more or less. Some of the talented are no more responsible for their ability to get to a much higher-than-average level of welfare from a given level of resources than the handicapped are for their handicaps; a distribution, for example, that leaves a handicapped person at a low level might also avoid allowing those

with talents for which they are not responsible to reap and hoard their rewards. Thus it might come closer than any alternative distribution to giving each person just what he deserves, which after all is relative to what others similarly deserving get. Rather, the information that a distribution would leave a handicapped person at a certain level of welfare tells us that circumstances obtain in which the theory assigns certain weights to considerations of equality of welfare and of responsibility in relation to other values. The weightings associated with these theoretically relevant circumstances may themselves come into conflict, if, for example, the only alternatives available will either leave someone below the minimum level of resources or leave a handicapped person below the minimum level of welfare. In such a case, further distinctions of circumstance and weight must be drawn; we should never assume the set of relevant differences between alternatives is closed. To require the theory to assign one weight to one value and another to another, once and for all, would be to render it gratuitously crude, to prevent it from ramifying into a set of second-order criteria that are satisfied to different degrees in different circumstances.

Finally, the counter-objection may be made on behalf of  $N^*$  that if the use of non-criterial information is not to enrich our understanding of how the given criteria bear on various alternatives, then such supposedly non-criterial information must in effect be functioning as an additional criterion.<sup>23</sup> If we assign different weights to our values in different circumstances, we must have reasons for doing so; the non-criterial information in terms of which we distinguish such circumstances just reflects such reasons. For example, the counter-objector might claim that falling below a certain level of welfare is a bad thing in itself.

However, this counter-objection does not succeed, on several grounds. At least part of the sense in which falling below a certain level of welfare is a bad thing is already captured by considerations of welfare. Moreover, an acceptable variant of the theory I have described might unweight considerations of equality of welfare when a more equal distribution would leave someone below an even lower level; decisions about how to allocate medical resources

---

23. Compare the argument that counter-examples to the sure – thing principle can be avoided by properly individuating possibilities; see John Broome, 'Rationality and the Sure-Thing Principle', Department of Economics, University of Bristol, Discussion Paper 84/158, September 1984. See also Amos Tversky, 'A Critique of Expected Utility Theory: Descriptive and Normative Considerations', 1975 pp. 163-73.

often seem to reflect such an unweighting.<sup>24</sup> The information that certain distributions would leave someone in a band between these two levels of welfare, where considerations of equality receive extra weight, hardly counts as a distinctive sort of reason for or against them. We need have no other reason for assigning different weights to our values in different circumstances than we would for assigning them one particular set of weights rather than another in all circumstances, namely, that by doing so we produce a better theory of the cases we are trying to explain. Furthermore, judgements about what ought to be done, all things considered, when values conflict, sometimes require us to draw arbitrary lines: consider how we would determine at what age someone is eligible to vote, or to marry without parental consent, or (assuming we think some abortions are permissible) during what period an abortion is permissible. Perhaps there is a presumption against drawing arbitrary lines, such that we sometimes determine that an earlier decision was mistaken rather than draw an arbitrary line between cases that would reconcile them. But the presumption is not irrebuttable; in theorizing we may sometimes draw arbitrary lines between cases so as to avoid attributing too many or too serious mistakes. (Recall my discussion of  $D^*$  above.) There is no reason to deprive ourselves of this theoretical flexibility and the information it requires.

We can resist  $N^*$ , then, on the grounds that a deliberator may reasonably require access to non-criterial information about the alternatives that  $N^*$  denies him. Thus impossibility results that follow from  $N^*$  do not threaten the existence of coherence functions.

Note that, having rejected condition  $N^*$ , we cannot proceed to reproduce impossibility results simply by treating non-criterial information as the source of additional rankings to which the relevant conditions apply, as non-criterial information is not directional in the way required for application of condition  $P^*$ . That is, that an alternative might be nearer to one extreme or the other of a ranking generated by non-criterial information would not in itself count for or against it; as explained in section 2 above, this is just what it is for information to be non-criterial.

---

24. See James Griffin, 'Equality: on Sen's Weak Equity Axiom', *Mind*, 1981, pp. 280-6, at p. 282.

## 6. The analogues of Roberts's single-profile conditions and supervenience

Roberts has obtained a single-profile impossibility result that depends both on a mild reformulation of the independence condition, which amounts merely to a requirement that cases alike in all respects be treated alike by the social welfare function, and on a potent reformulation of the condition of unrestricted domain.

Roberts shows how a reformulation of condition I can avoid incorporating neutrality with respect to non-welfare information by invoking a partitioning of the set of alternatives into equivalence classes such that all members of any such class have exactly the same non-welfare characteristics:

*L (Like Cases):* For any alternatives  $w$ ,  $x$ ,  $y$ , and  $z$ , if  $x$  and  $w$  are members of one equivalence class collecting alternatives with just the same non-welfare characteristics and  $y$  and  $z$  are members of another, and if the preferences of all individuals as between  $x$  and  $y$  are the same as their preferences as between  $w$  and  $z$ , then the preference of society as between  $x$  and  $y$  must be the same as its preference as between  $w$  and  $z$ .

If  $x$  and  $w$  are members of one such class and  $y$  and  $z$  are members of another, then any non-welfare information that might influence a decision between  $x$  and  $y$  will also apply to a decision between  $w$  and  $z$ , and vice versa. Thus, only welfare information could differentiate the two decisions to account for a decision in favour of  $x$  in one case and a decision in favour of  $z$  in the other.<sup>25</sup> Condition L says that if all welfare and non-welfare characteristics of two pairs of alternatives are the same, then the two issues should be resolved in the same way, or that like cases should be treated alike; this is just to require that judgements about social welfare supervene on judgements about individual welfare *and* all other information.

L's deliberative analogue is:

*L\* (Like Cases):* For any alternatives  $w$ ,  $x$ ,  $y$ , and  $z$ , if  $x$  and  $w$  are members of one equivalence class collecting alternatives with just the same non-criterial characteristics and  $y$  and  $z$  are members of another, and if the rankings by all criteria of  $x$  and  $y$  are the same as their rankings of  $w$  and  $z$ , then the ranking of  $x$  and  $y$  all things considered must be the same as the ranking of  $w$  and  $z$  all things considered.

---

25. See Roberts, *op. cit.*, at p. 443.

L\* says that if all criterial and non-criterial characteristics of two pairs of alternatives are the same, the two issues should be resolved in the same way; this is just to require that all-things – considered evaluations supervene on criterial *and* non-criterial information. This requirement cannot be disputed. It is a requirement on any reasonable theory that cases alike in *all* respects be treated alike, even if non-criterial differences alone may occasionally justify different treatment. We can not defuse an impossibility result for coherence functions by resisting L\*.

However, Roberts's result depends not only on a mild reformulation of independence, but also on a potent reformulation of unrestricted domain. Recall that multiple-profile condition U says that, for any set of alternatives and any set of individuals, the domain of a social welfare function includes all orderings of alternatives by individuals; the move to a single profile means that the domain of the social welfare function is *any one* ordering of alternatives by individuals, but *only one* (as Samuelson says). With this change, Arrow's result eludes us, as his proof of the emergence of a dictator depends on the ability U guarantees to hold the set of alternatives and the set of individuals constant and vary the orderings of the former by the latter.<sup>26</sup> As we have seen, a strong reformulation of independence as neutrality makes up for this weakening of the domain condition and permits impossibility results to be obtained for a single profile. Roberts provides an alternative route to an impossibility result for a single profile, which avoids neutrality and depends instead on a condition that uses the partitioning into equivalence classes to guarantee the richness of the single-profile domain:

*R (Richness):* For any ordered triple  $\langle X, Y, Z \rangle$  of equivalence classes, which collect alternatives with just the same non-welfare characteristics, and any ordered triple  $\langle A, B, C \rangle$  of complete specifications of an alternative's welfare characteristics, there exist three distinct alternatives characterized by A and membership in X, by B and membership in Y, and by C and membership in Z, respectively, for some conceptualization of welfare, or information base, equivalent to that of the welfare information actually realized.<sup>27</sup>

---

26. See Sen, *Collective Choice and Social Welfare*, pp. 43–4; see also Sen, *Choice, Welfare and Measurement*, p. 251 ff.

27. See Roberts, *op. cit.*, pp. 442–3.

The deliberative analogue of R is:

*R\** (*Richness*): For any ordered triple  $\langle X, Y, Z \rangle$  of equivalence classes, which collect alternatives with just the same non-criterial characteristics, and any ordered triple  $\langle A, B, C \rangle$  of complete specifications of criterial characteristics, there exist three distinct alternatives characterized by A and membership in X, by B and membership in Y, and by C and membership in Z, respectively, for some conceptualization of criteria equivalent to the actual (whatever this may be, which must be taken as a starting-point).

Intuitively, *R\** requires that the set of all alternatives be very rich, in that it contains alternatives with any combination of complete non-criterial characterization and complete criterial characterization we care to specify. Note that A, B, and C do not each specify information provided by one value; for example, A would not tell us about the justice of alternatives, something else about them, and so on. Rather, A might tell us that an alternative is ranked so by justice and so by kindness and so on. B would then give a different but also complete specification of criteria! information about an alternative derived from the same set of criteria.

However, while *L\** guarantees the supervenience of all-things-considered evaluations on information provided by specific values and all other information, by demanding such a rich set of alternatives *R\** fails to respect the supervenience of specific evaluative judgements on non-evaluative information. To see this, let the triple of equivalence classes be  $\langle P, P, P \rangle$  and the triple of complete specifications of criterial characteristics be  $\langle F, G, H \rangle$ , where F, G, and H are distinct and derive from evaluative characteristics. Then *R\** says that, for this choice of triples (which it entitles us to make), there are three distinct alternatives, all with non-criteria! specification P, but with different complete criterial specifications F, G, and H. That is, the complete evaluative specifications for alternatives with just the same non-evaluative characteristics differ. For complete evaluative specifications to differ, at least two values have to differ about them, even if all but one are indifferent between them. If two values differ about two alternatives, then even if one is indifferent at least one must itself distinguish between the alternatives. So, for the complete evaluative specification of two

alternatives to differ, at least one value must itself treat them differently. But if they are exactly the same in non-evaluative respects, this violates supervenience.<sup>28</sup>

Perhaps Roberts's result is preserved if  $R^*$  is altered to apply to triples of distinct equivalence classes. If so, however, we get a violation of supervenience by making two applications of condition  $R^*$ . For example, we might in the first case choose  $\langle P, R, S \rangle$  and  $\langle F, G, H \rangle$ , and in the second case choose  $\langle P, R, S \rangle$  and  $\langle G, H, F \rangle$ . Then consider the alternative characterized by membership in  $P$  and by  $F$ , and the alternative characterized by membership in  $P$  and by  $G$ . Again, we have alternatives with the same non-evaluative and different evaluative characteristics, which violates supervenience.

It might be suggested that condition  $R^*$  can be met in this case without violating supervenience, just by shifting from one conceptualization of some criterion to another between applications, for example, from a conception of equality of resources to one of equality of welfare.<sup>29</sup> The different conceptualization would then account for the different evaluative specification, without a violation of supervenience. However, the various permitted conceptualizations are supposed to be equivalent to that actually realized which must be taken as a starting-point. If in fact we give weight to considerations of both equality of resources and equality of welfare, and neither educes to the other, then the actual

---

28.  $F$ ,  $G$ , and  $H$  are analogous to instantiations of Roberts's  $b_1$ ,  $b_2$ , and  $b_3$ , which are the values for different alternatives of some function  $v$ , equivalent to the function that captures the welfare information that is actually realized, of one entire profile of preferences. Thus  $b_1$ ,  $b_2$ , and  $b_3$  each reflect the preferences of all the individuals included in the profile, and do not each correspond to a different individual.

An objection might be that all that is needed is for the two alternatives to differ with respect to some other value, though they are alike in non-evaluative respects; two or more values might form of circle of such distinctions between two alternatives with the same non-evaluative characteristics. But then all things considered evaluations delivered by the coherence function may be sensitive to evaluative distinctions which no non-evaluative distinctions underlie, which is surely an objectionable violation of supervenience, though pushed up to the next level. In fact I think the objection is wrong and there is an objectionable violation of supervenience to begin with.

The analogous complaint about condition  $R$  is that it requires someone to have different preferences with respect to alternatives that have just the same non-welfare characteristics. It may not worry some social choice theorists that as a result some preferences must be fundamentally arbitrary or meddlesome. However, if an epistemic justification is contemplated for democratic adherence to the recommendations of a single-profile social welfare function, perhaps such fundamental arbitrariness should be a cause for concern. That opinions about what ought to be done are fundamentally arbitrary may tend to undermine epistemic justifications for relying on them; such justifications might require opinions to be sensitive to the truth and might not hold irrespective of the content or accuracy of the opinions. Compare Arrow, pp. 85, 105.

29. This would be analogous to shifting from one  $v$  to another between two applications of Roberts's condition. See Roberts, *op. cit.*, pp. 342–3.

conceptualization involves two criteria and both must feature in any equivalent conceptualization. If we do not give weight to both, or one does reduce to the other, then any conceptualization equivalent to the actual must reflect this. We must either count both of them, or choose between them. Again, a coherence function represents a substantive theory about the relationships between various values; we must take as a starting-point, to theorize about, some definite conception of what those values have to say; if this is left entirely open, the theory is deprived of its substance. We have to know what equality actually does require, in various settled and hypothetical cases, in order to theorize about what its relationship is to other values. The theory we seek is about the values we actually have, not what they might have been. Thus, we can rule out variation in the conceptualization of criteria between applications as the explanation of different evaluative specifications. Even the altered condition  $R^*$  involves a violation of supervenience.

Since we are entitled by  $R^*$  to specify any criteria, we may specify the supervenient criteria, or values, that are of interest to us. Hence we may reject  $R^*$  on the grounds that it violates the requirement that evaluative descriptions supervene on non-evaluative descriptions, so that it cannot be a reasonable constraint to impose on deliberation about the relationships among conflicting values.

Thus the existence of coherence functions that takes as arguments rankings generated by values, constrained by supervenience, is not threatened by impossibility results that depends on  $R^*$ .

## **7. The analogue of multiple-profile condition U**

Recall that our digression into single-profile impossibility results was prompted by the sense that the independence condition was trivial as it stood and needed to be reformulated. The challenge to coherence accounts raised by single-profile impossibilities has not succeeded, but we are still left with the trivial sense in which  $I^*$  cannot be violated by coherence functions from supervenient criteria (if only because it is vacuous) and the original multiple-profile result. Now perhaps the analogous result is not a threat just because of the vacuousness of  $I^*$ . But suppose we admit, for the sake of argument, that since one way or

another,  $P^*$ ,  $D^*$ , and  $1^*$  hold, then either coherence functions do not exist or they do not satisfy  $U^*$ . Even if we admit this, however, still  $U^*$  does not hold, because it also violates supervenience.

$U^*$  says that the domain of the coherence function includes all orderings of alternatives by criteria. A proof of the emergence of a dictatorial criterion parallel to Arrow's would depend on  $U^*$  to guarantee the ability to hold the set of alternatives and the set of criteria constant and vary the orderings of the former by the latter. But this is just what the supervenience of evaluative criteria disallows. We can shift from one evaluative criterion to another, but a given evaluative criterion cannot vary independently of non-evaluative characteristics.

Should  $U^*$  be interpreted as requiring counter-evaluative counterfactual suppositions? We have already considered decisive objections to  $1^*$  interpreted counter-evaluatively, so even if  $U^*$  succeeds on this interpretation, we will not have shown that coherence functions do not exist. But let us consider counter-evaluative  $U^*$  anyway. Perhaps our values might have been other than what they are. The world might have been other than it is in many ways. We expect a scientific theory to tell us what to believe given the data we have about the way the world is; but we do not also expect it to tell us what we ought to have believed if the world had been different. If the world had been different, our theories about it might well have differed as a result. We do not hold it against a scientific theory or against scientific methodology that under counterfactual assumptions theories might be supported that are inconsistent with the theory that is in fact supported. Similarly, it is to misconstrue the role of a moral theory to expect it to tell us what we ought to have done if our values had been other than what they are; if they had been, again, we might well have needed a different, and inconsistent, theory.

Thus, on neither interpretation is  $U^*$  a reasonable condition to impose on coherence functions that take rankings generated by values as arguments and that represent the theories sought in deliberation; coherence is still possible.<sup>30</sup>

---

30. How essential is supervenience to avoiding impossibility results for coherence functions? Numerous domain restrictions that avoid Arrow's result have been discovered; for examples, see Sen, *Collective Choice and Social Welfare*, chapter 10.3. However, analogous domain restrictions for coherence functions

## 8. Summary and conclusion

To summarize the arguments that have led to this conclusion: the impossibility of social welfare functions that meet conditions P, D, I, and U has been shown by Arrow. If coherence functions can reasonably be required to meet formally analogous conditions, then their impossibility will follow. Coherence functions can reasonably be required to meet conditions P\* and D\*. With respect to 1\*: either coherence functions from supervenient criteria necessarily meet this condition, or the condition needs to be reformulated in a way that is appropriate for supervenient criteria. On the one hand, if the condition needs to be reformulated, we have two suggestions as to how to do so. The first would reformulate it as N\*, which requires neutrality with respect to non-criterial information. But this is not a reasonable requirement to impose on coherence functions: it would either render them gratuitously crude or would require us to revise our values to an extent that is incompatible with their discrete characters. The second suggestion would reformulate it so as merely to require that cases alike in all respects be treated alike. This requirement cannot be resisted. However, to obtain an impossibility result by means of this condition we need also to impose a reformulation of condition U\*. But the reformulation, R\*, violates supervenience, and is therefore not a reasonable condition to impose on coherence functions. If, on the other hand, coherence functions from supervenient criteria necessarily meet condition 1\* as originally formulated, then we must consider whether coherence functions can reasonably be required to meet condition U\* as originally formulated. But they cannot be, because such a condition would also violate supervenience.

Thus, when the criteria involved are values, impossibility results for coherence functions have not been found to obtain on the basis of the three sets of conditions we have considered: the combination of supervenience and non-neutrality makes it possible to avoid dictatorship. Coherence accounts have survived the challenge, and we may hope that deliberators are not seeking a chimera.

have a common fault: they would avoid impossibility results by stipulating that coherence functions do not apply to certain hard cases, namely, those involving the pattern of conflict illustrated in the text by the example of the three cases of an illness.

*St. Edmund Hall*  
*Oxford*

S. L. HURLEY