A welfarist critique of social choice theory

Aki Lehtinen

Journal of Theoretical Politics 2011 23: 359
DOI: 10.1177/0951629811411753

Article

JOURNAL OF THEORETICAL POLITICS

A welfarist critique of social choice theory

Journal of Theoretical Politics
23(3) 359–381

DOI:10.1177/0951629811411753
jtp.sagepub.com
SAGE

Aki Lehtinen
University of Helsinki, Finland

Abstract
This paper reconsiders the discussion on ordinal utilities versus preference intensities in voting theory. It is shown by way of an example that arguments concerning observability and risk-attitudes that have been presented in favour of Arrow's Independence of Irrelevant Alternatives (IIA), and against utilitarian evaluation, fail due to strategic voting. The failure of these two arguments is then used to justify utilitarian evaluation of outcomes in voting. Given a utilitarian viewpoint, it is then argued that strategy-proofness is not normatively acceptable. Social choice theory is criticised not just by showing that some of its most important conditions are not normatively acceptable, but also by showing that the very idea of imposing condition on social choice function under the assumption of sincere behaviour does not make much sense because satisfying a condition does not quarantee that a voting rule actually has the properties that the condition confers to it under sincere behaviour. IIA, the binary intensity IIA, and monotonicity are used as illustrations of this phenomenon.

Keywords
IIA; preference intensity; strategic voting; strategy-proofness; utilitarian winner

1. Introduction

Arrow's theorem and the Gibbard–Satterthwaite theorem are commonly taken to be the most fundamental results in areas of social choice theory that deal with voting. In this paper, I argue that these theorems have very little normative relevance because the conditions upon which they are based are not normatively acceptable.

The normative and descriptive relevance of preference intensities and the normative validity of Kenneth Arrow's Independence of Irrelevant Alternatives (IIA) have been

Corresponding author:
Aki Lehtinen, Department of Political and Economic Studies, practical philosophy unit, PO Box 24, 00014 University of Helsinki, Finland
Email: aki.lehtinen@helsinki.fi

under debate for decades in the context of social choice theory. IIA can be defined as follows. Let $C(S)$ denote a choice made by society in voting from a set of alternatives $S \subset X$. Let $\mathbf{p}$ and $\mathbf{p}'$ denote profiles of individual preferences: $\mathbf{p}$ assigns a preference ordering $\succ_i$ for each voter $i \in I : \mathbf{p} = (\succ_1, \succ_2, ..., \succ_N)$. Let $\mathbf{p}_{|Y}$ denote the restriction of the profile $\mathbf{p}$ to the subset Y of X. Let $C(\succ, S)$ denote the social choice from profile $\mathbf{p}$ on S.

Independence of Irrelevant Alternatives: For all $x, y \in S$ and all individuals,

$$\mathbf{p}_{|S} = \mathbf{p}'_{|S} \rightarrow C(\succ, S) = C(\succ', S). \tag{1}$$

In other words, if the two profiles $\mathbf{p}$ and $\mathbf{p}'$ rank each pair of alternatives in the same way, then the social choice should be the same.

Preference intensities are taken into account in various models of strategic voting that describe the behaviour of voters under uncertainty.[1] At the same time, however, the use of a utilitarian welfare function in evaluating voting rules is rare. I will take it as given that intensities of preference are intrinsically relevant for evaluating voting outcomes normatively. I believe that voting theorists, including Arrow (1987), agree with this judgment, and I will thus not attempt to justify it. The reason for this discrepancy between positive and normative approaches is thus that Arrow and others have presented arguments for why one should not use the utilitarian welfare function.

Traditional criticisms of preference intensities can be formulated in terms of two arguments for IIA. The *observability argument* states that since it is possible to observe preference orderings, but not preference intensities or interpersonal comparisons of utilities, allowable information must be restricted to preferences for pairs of alternatives, and this is what IIA does. The *epistemological-moral argument* against preference intensities and for IIA states that von Neumann–Morgenstern (vNM) utilities should not be used in social-welfare judgements because they 'reflect only individuals' attitudes towards gambling' (Arrow, 1951: 9–11).[2] The idea here is that vNM utilities are not appropriate in this context because they inevitably incorporate attitudes towards risk.

Although Arrow may not have introduced IIA in order to preclude strategic voting,[3] this seems to be the most important consideration for those who continue to think that IIA is normatively acceptable. The basis for such views derives either from intuitive considerations (Plott, 1976; Vickrey, 1960) or from various proofs that link IIA and strategy-proofness in some way.[4] The *strategic-voting argument* thus states that strategic voting is to be avoided, and a voting procedure that satisfies IIA precludes it.[5]

I will respond to these points as follows. I will show with a simple example of strategic voting under amendment agendas that if Arrow's assumption that voters are sincere is dropped, *none* of the properties that are commonly attributed to social choice functions[6] that satisfy IIA are actually found in the actual voting procedure that the social choice function was supposed to represent: if there is strategic voting, third alternatives affect the choice between a pair, and preference intensities as well as attitudes towards risk affect the outcome.

I will thus not attempt to show that preference intensities are observable, or that we have particularly precise information on interpersonal comparisons. I will rather establish that observability cannot be used as an argument against preference intensities in evaluating voting rules. As an argument for IIA, the epistemological-moral argument suffers from a similar shortcoming: voting choices reflect attitudes towards risk also under voting rules that satisfy IIA. Whereas this fact shows that the requirement of not taking attitudes towards risk into account cannot be satisfied by any voting rule, it does not disarm the normative force of the argument. However, if there is a way of modelling voting in such a way that the utilities themselves do not contain risk-attitude information, even though the idea that risk attitudes should not affect voting results is prima facie plausible, if the typical aggregate-level consequences of voting are better when voters, choices are affected by risk attitudes than when they are not, even the normative force of this argument is discharged.

The example simultaneously shows that strategic voting may occur in a voting procedure that satisfies IIA. Although this point is no longer new given that it should have been known ever since Enelow (1981) presented his model of strategic voting under amendment agendas, it may be worthwhile analysing it more thoroughly in view of the fact that even though Saari (2008) mentions it in one sentence, Munger (2009) remains unconvinced. Saari's observation may fail to convince those who think that IIA is to be justified on grounds that have to do with strategic voting because they think that IIA has something to do with how people behave when they vote. Saari's observation merely establishes that IIA does not guarantee strategy-proofness, but it does not show that there is anything wrong with strategy-proofness itself from the normative point of view. Munger thus argues that even though IIA does not guarantee strategy-proofness, voting rules that satisfy IIA are better than those that violate it because they are less susceptible to strategic voting.

I will follow Saari by defining IIA in such a way that it only concerns how preferences or ballots are aggregated, and it has nothing to do with how people behave when they vote. I should not be blamed for being partial against IIA in adopting this way of defining IIA because if it were interpreted in such a way that it takes behavioural assumptions into account, this would be even worse for its normative acceptability because it would then preclude the beneficial consequences that strategic voting typically entails. Munger's argument thus also fails because, as the example shows, strategy-proofness is not normatively desirable since strategic voting yields better outcomes in utilitarian terms than sincere voting.

Although rebutting the observability and the epistemological-moral argument is important in completing the list of failed arguments for IIA, the main purpose of these rebuttals is to establish the failure of the main arguments against utilitarian evaluation of outcomes in voting theory, and thereby justify the normative criticism of strategy-proofness. Unlike some previous non-welfarist arguments that challenge strategy-proofness (van Hees and Dowding, 2008), my critique fundamentally depends on the idea of evaluating the *consequences* of strategic voting in utilitarian terms.

This paper merely gives an illustrative example, but the result is very general: it has been derived under so many commonly used voting rules that if one were now to find a voting rule in which strategic voting is harmful, it would merely be an argument against that particular rule rather than the idea that strategic voting is typically beneficial. Here I will draw on the simulation results in Lehtinen (2007a, 2007b, 2008, 2010), which suggest that utilitarian efficiency (the frequency with which the alternative with the highest sum of utility is selected) is higher if voters engage in strategic behaviour than if they always vote sincerely. Strategic voting is thus unambiguously beneficial under the utilitarian evaluation of outcomes because it typically increases utilitarian efficiency (or average utility) as compared to sincere voting. Furthermore, the reason for this result is general. Voter behaviour depends on preference intensities when the voting is strategic but not when it is sincere: many strategic votes for the utilitarian winner are *counterbalanced* by few such votes against it. I henceforth refer to these models collectively as the 'counterbalancing model'. Voters' choices thus reflect preference intensities, but only in the case of strategic voting. They illustrate how all of the voting rules studied take intensity information into account, and this has beneficial aggregate-level consequences.

IIA, or the very idea of formulating social choice problems in terms of functions that take preference orderings as arguments, may also be taken to be an expression of welfarism. Welfarism has some appeal in voting theory if only because the very purpose of the theory is to find rules that best satisfy preferences (cf. Arrow, 1997). For the purposes of this paper I will take welfarism as given, and the discussion will focus on whether or not preference intensities should be taken into account in normative evaluations of voting schemes.

Before setting off, it may be worth explaining the nature of this paper. Even though I am challenging the normative acceptability of some important conditions, my main goal is not to find an alternative set of conditions that could be taken to characterize desirable properties of voting rules. The point is rather to show that the very idea of imposing conditions on social choice functions under the assumption of sincere voting is problematic.

The general problem is that the question of whether or not any given condition is satisfied by a social choice function (or welfare function or voting procedure) under the assumption of sincere voting is trivial, because the properties that the condition confers on the social choice function under sincere voting are no longer guaranteed to hold if voters engage in strategizing. This problem is ubiquitous in social choice theory because we already know that strategy-proofness can virtually never be guaranteed. The problem concerns not just IIA but also just about any condition that has ever been proposed in social choice theory. As further examples of the problem, I also show that if people vote strategically, although the Borda count satisfies Saari's (1998, 2001, 2003a, 2003b) alternative to IIA, the *binary intensity IIA*, and monotonicity, it does not necessarily provide us with correct binary intensity information and it is possible that increasing an alternative's position in a voter's preference ranking may lead to its defeat. The Borda count thus does not escape my critique either. In general, my point is not to criticize or justify any particular voting rule, not even the 'utilitarian' rule (Hillinger, 2005) or range voting (Smith, 2000).

The structure of this paper is as follows. Sections 2 and 3 present the epistemological-moral and the observability argument, respectively. Section 4 presents Enelow's model of strategic voting under amendment agendas, and Section 5 discusses how the counterbalancing models escape the epistemological-moral argument. Enelow's model is then applied in Section 6 to show that IIA guarantees none of the properties that it has been claimed to have. Section 7 provides another example in order to show that satisfying the intensity IIA and monotonicity does not really guarantee these properties if voters are strategic.

2. The epistemological-moral argument

Arrow and Rawls first presented what I call the epistemological-moral argument as a criticism of John Harsanyi's position. It posits that vNM utilities should not be used in social-welfare judgements because they inevitably contain morally irrelevant information on attitudes towards risk. The *moral* aspect is that attitudes towards risk are irrelevant to social-welfare judgments and they should therefore not be taken into account, and the *epistemological* aspect is that vNM utility functions can only be constructed from choices involving risk. [7] Hence, attitudes towards risk inevitably affect social-welfare judgements if these judgements are based on vNM utilities.

Harsanyi has persistently argued that vNM utility functions may be used for social-welfare judgements: they express a willingness to take risks in order to obtain some particular alternative (Harsanyi, 1987). Hence, they express the relative intensity with which a person prefers one alternative to another (see also Harsanyi (1978, 1979) and Ng (1999)).

Harsanyi (1992: 682–684) claims that Arrow and Rawls confuse 'process utility' and 'outcome utility' (see also Harsanyi, 1993). Process utility, or 'utility from gambling', refers to enjoyment from playing a game that involves risk, whereas outcome utility relates to the prizes one may obtain. Harsanyi is right in that the reduction of the compound-lotteries axiom precludes process utilities and thereby 'utility from gambling'. The vNM theory thus rules out attitudes towards enjoyment from gambling by assumption. Harsanyi is also right in pointing out that outcome utilities are ethically important. His arguments could be used to account for why we think preference intensities are morally relevant. However, the problem with his notion about process utility and outcome utility is that it does not really provide a response to the criticism: attitudes towards process utilities are not what a carefully stated epistemological-moral argument should be about. Arrow (1973b), for example, suggests that vNM utilities incorporate attitudes towards risk. The epistemological-moral argument also concerns attitudes towards risk that are related to voters' willingness to engage in strategic behaviour, not just attitudes towards enjoyment from gambling, and these attitudes are *also* irrelevant to social-welfare judgements.[8]

Arrow's and Rawls's position is buttressed by a well-known epistemological consideration in decision theory: standard expected utility theory does not provide any way of distinguishing between the psychological sensations of diminishing marginal utility (or diminishing the intensity of satisfaction) and risk aversion if all we are given are a person's *choices* under uncertainty. Indeed, Harsanyi (1992) admits this. According to the standard account of expected utility, preferences are the primitive concept in the theory, and they are defined over lotteries rather than final outcomes. Since preferences for lotteries are ordinal, there is a sense in which vNM theory is not a cardinal theory at all (Sen, 1976; Weymark, 2005), even though a cardinal preference schedule can be constructed using the so-called reference lottery technique (see, e.g., Hirshleifer and Riley, 1992), this requires assumptions that are not included in the theory; it is only by making assumptions concerning risk attitudes that we may interpret choices under risk as reflecting intensity information.

However, we may well make such assumptions, and then choices under risk may reasonably be taken to *reflect* preference intensities, just as Harsanyi claims, but this does not change the fact that attitudes towards risk *also* affect these choices. Hence, whereas vNM utilities incorporate ethically relevant information concerning preference intensities, they also incorporate ethically irrelevant information concerning attitudes towards risk.

If preference intensities exist in the first place, then Harsanyi has successfully shown that choices under uncertainty reflect them, and that they are morally relevant. At the very least, it seems natural to assume that both intensities and risk attitudes affect choices under uncertainty (cf. Broome, 2008a). Note that this is different from claiming that such choices provide us with reliable information on intensities: such information is always tainted with information concerning risk attitudes. The epistemological-moral argument thus remains valid because vNM utilities inevitably reflect morally irrelevant attitudes towards risk. However, it could be used against preference intensities in social choice theory only if it is possible to collect reliable information on ordinal utilities that *do not* reflect attitudes towards risk. I will show in Section 4 that this is not possible.

3. The observability argument

Those who have opposed the use of preference intensities and vNM utilities in social-welfare judgements have based their criticism on epistemological considerations. Here are Arrow's reasons for not incorporating preference intensities into social-choice theory:

> The oldest critique of social choice theory ... is that it disregards intensity of preference. Even with two alternatives, it would be argued that a majority with weak preferences should not necessarily prevail against a minority with strong feelings ... The problem in accepting this criticism is that of making it operational. Theoretically, is there any meaning to the interpersonal comparison of preference intensities? Practically, is there any way of measuring them, that is, is there any form of individual behavior from which the interpersonal comparisons can be inferred? (Arrow, 1977)

Arrow introduced IIA in order to impose an observational requirement on social choices. 'Modern economic theory has insisted on the ordinal concept of utility; that is, only orderings can be observed, and therefore no measurement of utility independent of these orderings has any significance [...] The condition of IIA extends the requirement of observability one step farther'(Arrow, 1983[1967]: 75–76).[9] His idea was that the available information had to be restricted to ordinal utilities because preference orderings were observable but intensity was not. Indeed, he made it perfectly clear that cardinal utilities (preference intensities) would be important for social choice and welfare if we could observe them (Arrow, 1987).

Arrow (1973a) argued thus: 'In a voting context, the ordinalist-cardinalist controversy becomes irrelevant, for voting is intrinsically an ordinal comparison and no more'. Strasnick (1976: 243) formulates the difficulty of observing preference intensity in a voting context as follows: 'There is no sense in which the magnitude or degree or intensity of a choice is observable in the choice itself'.

This, however, does not mean that voters' choices are unaffected by preference intensities. Example in which the outcomes *depend on preference intensities* even under a voting rule in which voters may express a preference directly only for pairs of alternatives (the majority rule with an amendment agenda) are given in Sections 6 and 7. They show that voting is intrinsically an ordinal phenomenon only in the sense that voters can merely state whether one alternative is better than another in pair-wise contests. However, if voters engage in strategic behaviour, their choices inevitably reflect preference intensities, and they *affect the outcomes* even under a rule that seemingly collects only ordinal information. In order to elaborate on these examples I will present a rudimentary version of a model of strategic voting under the majority rule (Enelow, 1981; Lehtinen, 2007b), and discuss the status of utilities in Lehtinen's version of this model.

4. Voting is not intrinsically an ordinal comparison

Let $X = \{x, y, z\}$ denote a set of available alternatives and U^i voter i's utility function. Table 1 shows the possible preference orderings.

Table 1. Voter types and utilities

type of voter						
t_1	t_2	t_3	t_4	t_5	t_6	U^i
x	y	z	x	y	z	Max
y	z	x	z	x	y	Med
z	x	y	y	z	x	Min

Alternatives are put in a sequence of pair-wise majority comparisons in *an amendment agenda* which is depicted in Figure 1.

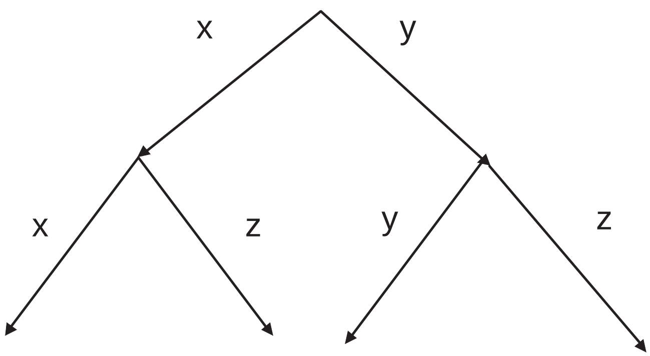

Figure 1. An amendment agenda

Two alternatives, x and y, are put to a majority vote against each other in the first round of voting. The winner of this first contest is then put to vote against the third alternative z in the second round.

Voter i's subjective probability that a given alternative j beats another alternative k $(j, k \in X)$ in a pair-wise second-round contest is denoted p^i_{jk}. In the first round of voting, voters choose a branch in the voting tree by comparing expected utilities for lotteries $(x, z; p^i_{xz}, 1 - p^i_{xz})$ and $(y, z; p^i_{yz}, 1 - p^i_{yz})$. Note that merely formulating the voters' choice situation under incomplete information shows that they are making a choice not between the pair $\{x, y\}$, but rather between two lotteries that *also involve the third alternative z*. It follows immediately that their 'choice between x and y' in the first round *may* reflect preference intensities for the various alternatives. Expected-utility expressions need to be formulated in order to show this.

Maximizing expected utility implies giving one's vote to the branch in the voting tree that has the greatest expected utility. A voter will vote for x rather than for y if

$$p^i_{xz} \cdot U^i(x) + (1 - p^i_{xz}) \cdot U^i(z) \geq p^i_{yz} \cdot U^i(y) + (1 - p^i_{yz}) \cdot U^i(z). \tag{2}$$

Consider now voter types one and four. Both prefer x to y, but the preferences of type-four voters are ordinally more intensive because they separate the preferences between these alternatives with z by preferring x to z to y, whereas type-one voters prefer x to y to z.[10] Type-four voters have a dominant strategy to vote sincerely for x.

The counterbalancing models use utility numbers that are generated randomly from the [0,1] interval. If a utility function U can describe voters' behaviour, so can positive affine transformations $V = a + bU$. Supposing that each voter has three such utility numbers (Max, Med, Min) for the three alternatives, voters' behaviour can always also be described as if their utilities were normalized as Max $= 1$, and Min $= 0$ and $v_i = \frac{(\text{Med}-\text{Min})}{(\text{Max}-\text{Min})}$.[11] Then v_i provides a natural measure for voter i's intrapersonal *intensity* of preference: if v_i is close to one, voter i considers the second-best alternative almost as good as the best one, and if it is close to zero, he or she considers it almost as bad as the worst one. The traditional notion of intensity is formalized as cardinality of utility functions. Preference intensities can be expressed if the utility functions allow for making judgments concerning differences in utility. It must be meaningful to say, for example, that i prefers x to y more than z to w: $U_i(x) - U_i(y) > U_i(z) - U_i(w)$. It is clear that such judgments could be expressed with utilities in the counterbalancing models if they were to use four or more alternatives. The v_i parameters thus do model preference intensity.

Applying such a utility normalization to a type-one voter yields

$$p^i_{xz} \cdot 1 + (1 - p^i_{xz}) \cdot 0 \geq p^i_{yz} \cdot v^i_1 + (1 - p^i_{yz}) \cdot 0. \tag{3}$$

Type-one voters will thus vote strategically for y if:

$$v^i_1 > \frac{p^i_{xz}}{p^i_{yz}}. \tag{4}$$

When they do, they are effectively expressing a cardinally strong intensity for x and y over z, and a cardinally weak intensity between x and y: type-one voters who have a strong intrapersonal preference intensity for y (high v_i) are more likely to vote strategically for y than those with a weak intensity. If $p^i_{xz} < p^i_{yz}$, there will be some value of v^i at which type-one voters will vote strategically. Hence, they express their preference intensity between x and y by deciding whether or not to vote strategically. In contrast, type-four voters never vote for y in the first round, and thereby reveal a strong intensity of preference for x over y. Voters thus express their ordinal and cardinal intensities under agenda voting, but they do this only in a probabilistic sense. Note that voters' choices depend on their risk attitudes because their choices depend on their beliefs. However, if beliefs are kept fixed and the preference intensities are changed by changing the v^i parameter, different behaviour will ensue. This shows that voters' behaviour depends on their preference intensities.

5. The nature of utility in the counterbalancing models

Given that I am arguing for a utilitarian evaluation of outcomes in voting theory, it would seem natural to take Harsanyi's (1953, 1955, 1977) theorems as a decision-theoretic justification for a utilitarian position. Harsanyi claims that the theorems show that vNM utilities represent preference intensities, and that they can be used to provide an argument for utilitarianism. I do not draw on these theorems because I fully accept the criticism that Harsanyi's utilitarianism is 'utilitarianism in name only': [12] the theorems do not really provide an argument. The essence of this critique is that because utilities represent preferences, and the representations are not unique, it is arbitrary to use vNM utilities for this representational purpose: a different utility transformation does not yield utilitarian welfare functions (Roemer, 2008). The important question is whether the fact that Harsanyi's argument for utilitarianism fails really implies that utilitarianism is untenable. This would be the implication if vNM utilities were the only possible way of conceptualising the notion of utility. It is indeed surprising how unanimously the connection between vNM utilities and utilitarianism seems to be accepted in the discussion on Harsanyi's theorem, given that many scholars would have preferred von Neumann and Morgenstern to have called their utility by some different name because of the possibility of confusing it with utilitarianism. Furthermore, there seems to be no good reason why utilitarianism needs a behaviourist foundation in people's choices in the first place, let alone in voting theory in which behaviourist arguments fail – at least if my argument in the next section is valid.[13]

Although the behavioural equations in Enelow's (1981) and Lehtinen's (2007b) models are identical, there is an important difference in how the utilities are to be interpreted. Given that Enelow disavows any connection with utilitarianism, the numbers used in his model are best understood as vNM utilities. Interpersonal comparisons of utility are meaningless with vNM utilities because they are supposedly constructed according to a procedure, the reference lottery technique, that only involves one person at a time. Weymark (2005) argues that although consistent behaviour could conveniently be described in terms of vNM utility functions, there is no particular reason to use this class of transformations for describing people's welfare for the purpose of making welfare judgments. However, voters have just one set of utilities in the counterbalancing models. It would be unnatural to use different utility transformations for describing the behaviour of voters and for describing their preference intensities for normative purposes. As far as I can see, the only reason for refusing to use the same transformations for these two purposes is that such an assumption implies a commitment to a particular interpersonal comparison.

More important, however, is the fact that the counterbalancing model is not an attempt to construct cardinal utility functions from axioms or behaviour. The idea is rather that the intensity information is already assumed to be incorporated into the utilities, and the zero-one transformation is merely used for pointing out how preference intensity could be expressed in voters' choices.

Interpersonal utility comparisons are made in these models under the assumption that the utility numbers are unique and fully comparable. It would be misleading to call such numbers 'vNM utilities', even though there is no particular reason why voters would violate the vNM conditions. The utility numbers are best understood as primitive: they describe voters' preference orderings and the intensity of preference. On account of the fact that beliefs are determined in a separate account (by signal extraction) the utilities are linear in the probabilities in equation (2) (just like vNM utilities), not because voters are assumed to satisfy the vNM axioms but because they are assumed to maximize expected utility in a literal sense: they are assumed to engage in the mental operation of *weighing the utilities of outcomes with the probabilities*. Saying that voters' utility numbers can be described as vNM utilities implies no such commitment: it merely implies that their behaviour complies with the axioms necessary for representation as vNM utilities.

The Sen–Weymark criticism thus does not concern the utilities in the counterbalancing models because they are unique by assumption, and they are primitives that are taken to describe preference intensities by assumption. The real question is thus whether this construction is acceptable.

The behavioural parts of the counterbalancing models should be acceptable because they are identical to earlier, already accepted accounts in voting theory. The tricky questions concern the way in which preference intensities and interpersonal comparisons are conceptualized. I will not say much about interpersonal comparisons here because the counterbalancing models already provide an account of why it is legitimate to make such comparisons in this model: the main result that strategic voting increases utilitarian efficiency is highly robust with respect to different interpersonal comparisons (Lehtinen, 2007b, 2008). This takes care of the epistemic part, not by showing how to obtain the relevant information but by showing that we do not need it: even though we will never know what would correspond to voters' 'real' utility scales, this does not matter because the results of the model hold under *all* normatively acceptable choices.

This argument presupposes the idea that there are limits to how large the acceptable interpersonal differences in utility may be. The purpose of voting theory is to evaluate the functioning of various voting rules. It is not normatively adequate to allow the assumption in such exercises that the difference between the minimum and the maximum utility of some single individual is, say, a thousand times more than that of others. If this were the case, and the sum of utility was used as a welfare function, it would essentially mean that the voting rule was being evaluated in a way that depended only on one individual's preferences. One would then be wondering why voting is used to begin with, given that social welfare is essentially based on one person's preferences. One reason why one individual has one vote under most rules is that each individual's voting choice is considered equally important, and each individual's utility is taken to carry at least roughly equal weight in the welfare function (see also Hammond, 1987). The one-man-one-vote principle may thus be taken to be implicitly based on rough equality of utility scales (cf. Mackie, 2003).

6. Amendment agendas do not have the properties IIA was supposed to provide: an example

It used to be common to distinguish between different aspects of the IIA condition.[14] The *independence* (or irrelevance) *aspect* refers to the fact that the social ordering between any two alternatives must depend only on individual preferences for these and not for other irrelevant alternatives. The *ordering aspect* requires that the social ordering (or choice) of any two alternatives must be based only on individual orderings of these alternatives and on nothing else. This aspect used to be taken to rule out preference intensities.[15]

Table 2. Example 1

A	B	C
y (1)	y (1)	x (1)
x (0.9)	x (0.9)	z (0.5)
z (0)	z (0)	y (0)

It is generally acknowledged that if relative intensities of preferences are somehow available then the ordering aspect of IIA need not be accepted. Furthermore, the 'irrelevant' alternatives are not, strictly speaking, irrelevant.[16] IIA does not distinguish between alternatives that are not even included in the set of available alternatives and those that belong to it but are not under explicit consideration at a given stage of voting. The truly irrelevant alternatives belong to the former set (cf. Bordes and Tideman, 1991; Hansson, 1973).

Consider an example of voting under an amendment agenda (depicted in Table 2).[17] Assume that all three voters have identical beliefs such that $p_{yz} = 0.7$ and $p_{xz} = 0.9$. Voters A and B are of type five. They will vote strategically for x in the first round because $v_5 < \frac{p_{yz}}{p_{xz}}$ is false ($0.9 \nless \frac{0.7}{0.9} = 0.7778$). Voter C has a weakly dominant strategy to vote for x in the first round. Thus, x is the outcome if the voters maximize expected utility because it beats y in the first round and z in the second round. The utilitarian winner x is chosen if they maximize expected utility but the Condorcet winner y is chosen if they vote sincerely. A Condorcet winner is thus not necessarily chosen under the majority rule.

Let $n(j \succ k) = \sum_{i=1}^{N} (j \succ_i k)$ denote the number of voters who prefer alternative j to k. The underlying social choice function, the method of majority decision, is defined by

$$\forall x, y \in S : C(S) = x \leftrightarrow \forall y \in S : n(x \succ y) > n(x \prec y). \tag{5}$$

A Condorcet winner (CW) is defined by

$$CW = \{x | x \in S : \forall y \in S : n(x \succ y) > n(x \prec y)\}. \tag{6}$$

Since the method of majority decision is defined in terms of the preferences rather than the expressed ballots, it declares y as the Condorcet winner and as the alternative that is selected. This is not what happens, however, if voters engage in strategic voting.

Let us now consider the observability argument. It only makes sense if IIA guaranteed that we may observe the *real* rather than the expressed preferences. In the example IIA is satisfied but the voting rule provided the ordinal information that a majority of voters prefer x to y – which is false as a statement about their real preferences. Ordinal utility is not observable either in the sense that the selected alternative need not be the Condorcet winner under the majority rule and amendment agendas. The sum-of-utility criterion has been criticized for not being observable (e.g., Arrow, 1973b). Preference orderings would be observable if the Condorcet winner were always selected under the majority rule, but this is not the case. The possibility of strategic voting undermines the observability argument. Therefore, preference orderings are not observable either, and observability is not a valid argument for ordinal utility and against intensities in a voting context.

The claim that preference orderings are scientifically respectable because they can be observed is invalid against intensities *in voting theory* even though it may have some weight in other contexts. I take it that the observability argument has not been taken seriously for quite a while. Blin and Satterthwaite (1977), for example, point out that if we knew the preferences with certainty, the need for a legislative body would vanish because preferences could be aggregated directly.

It would, of course, be easier to collect information on preference orderings than on intensities by means other than voting. We could, for example, simply ask the voters about their preference orderings. The problem with any procedure other than voting, however, is that insofar as the results are used for making decisions, individuals have an incentive to misrepresent their preferences. If, on the other hand, the results are not used for making decisions, voters, particularly representatives in parliaments' have an incentive to misrepresent their preferences in order to give signals to their constituencies. Collecting information on preference orderings is thus easier than collecting information on preference intensities, but it is ultimately not possible to obtain fully reliable information on either.

The example also shows that the epistemological-moral argument is not tenable either because attitudes towards risk and preference intensities *inevitably* affect voting choices if voters maximize expected utility under incomplete information. Attitudes towards risk always affect voters' choices simply because their behaviour depends on their beliefs, and their beliefs depend on their risk attitudes. As explained in detail in Lehtinen (2006), counterbalancing models formalize attitudes towards risk in terms of voters' degree of confidence in perturbed signals concerning the preference profile. Their degrees of confidence thus have an effect on the exact numerical values of their beliefs, and thereby affect their propensity to engage in strategic voting. Since the beliefs are determined separately from the utilities in this model, the utilities themselves do not contain any information on risk attitudes. Even though risk attitudes affect voters' behaviour, the utilities are untainted in the right way. I have thus provided an acceptable solution to the epistemological-moral argument: since the utilities themselves do not contain risk attitudes, whether or not they should affect voting outcomes becomes a question that could be answered on the aggregate level.

The question is whether we would prefer to live in a world in which risk attitudes affect the outcomes (through strategic voting) or in one in which they do not. I would not prefer to live in a counterfactual world in which risk attitudes do not affect outcomes because voters have complete information or are unable to vote strategically for some mysterious reason. The reason for this is already provided in the counterbalancing models (Lehtinen, 2007a, 2007b, 2008): in utilitarian terms, the current world in which voters live under uncertainty and engage in strategic voting is better.

It is not possible in this paper to prove the general claim that intensities will affect the results under all voting rules. However, it is clear that insofar as an expected-utility model can be formulated for any voting rule, it can be shown that preference intensities will affect the outcomes under it. It follows that if the epistemological-moral argument is to be effective against using intensities in voting theory, one has to deny that voting is characterized by decision-making under uncertainty. Surely, however, nobody is willing to argue that voters have complete information on other voters' preferences in an electorate of dozens, thousands or millions. Real-world voting is clearly characterized

Table 3. Example 2

1	2	3	4	5	6	7
y (1)	y (1)	y (1)	z (1)	z (1)	z (1)	x (1)
x (0.9)	x (0.9)	x (0.9)	y (0.2)	x (0.9)	x (0.9)	z (0.9)
z (0)	z (0)	z (0)	x (0)	y (0)	y (0)	y (0)

by decision-making under uncertainty, as Coleman (1966) argued long ago. A reasonable voting model should explicitly take this into account rather than circumventing the problem by using only ordinal utilities.

7. Another example

Assume that the preferences of seven voters can be described as in Table 3. The zero-one normalization is again used merely because it facilitates recognition of the role of intensities in the example.

The numbers in parentheses denote voters' utilities. The sums of utilities are $\Sigma_i U_i(x) = 5 \times 0.9 + 1 = 5.5$, $\Sigma_i U_i(y) = 4 \times 1 + 0.2 = 4.2$ and $\Sigma_i U_i(z) = 3 \times 1 + 0.2 = 3.9$. x is the utilitarian winner and z the worst outcome in utilitarian terms.

If all voters vote sincerely under an amendment agenda, y will beat x in the first round by four votes against three, and z then beats y in the second round by four votes against three, and the worst alternative in utilitarian terms emerges as the final outcome.

Let us now see what would happen if voters maximize expected utility under incomplete information. Assume that all three voters have identical beliefs such that $p_{yz} = 0.3$ and $p_{xz} = 0.8$. A voter gives a vote to x in the first round if the expected utility (EU) for x is higher than that for y. Voters 1, 2 and 3 are of type five. They vote strategically for x in the first round because $EU(x) = p_{xz}U(x) + (1 - p_{xz})U(z) = 0.8 \times 0.9 + (1 - 0.8) \times 0 = 0.72$ is larger than $EU(y) = p_{yz}U(y) + (1 - p_{yz})U(z) = 0.3 \times 1 + (1 - 0.3) \times 0 = 0.3$.

Similarly, the expected utilities of voter 4 are $EU(x) = 0.2$ and $EU(y) = 0.76$, and the expected utilities of voters 5 and 6 are $EU(x) = 0.92$ and $EU(y) = 0.7$. Voter 4 thus votes sincerely for y, and voters 5 and 6 vote sincerely for x in the first round. Voter 7 has a weakly dominant strategy to vote for x in the first round. Thus, x is the outcome if the voters maximize expected utility because it beats y (6–1) in the first round and z (4–3) in the second round. The utilitarian winner x is chosen if they maximize expected utility but z is chosen if they vote sincerely.

Voters 5 and 6 might wish to counteract this result by voting strategically for y. However, the logic of counterbalancing implies that this is unlikely because they would have to believe that y has virtually no chance against z, and that, simultaneously, x is almost sure to beat z. Even if they thought that x beats z with certainty ($p_{xz} = 1$), they would vote sincerely for x because $EU(x) = 1 \times 0.9 + (1 - 1) \times 1 = 0.9 > 0.7$. To vote strategically for y, they would also have to believe that $p_{yz} < 0.1$ ($EU(y) = 0.1 \times 0 + (1 - 0.1) \times 1 = 0.9$).

Note that voters 1, 2 and 3 would continue to vote strategically for x even if they had much less confidence in the chances of x against z in the second round. Keeping all the other parameters fixed, they vote strategically if $p_{xz} > 0.34$. ($EU(x) = 0.34 \times 0.9 + (1 - 0.8) \times 0 = 0.306 > 0.3$). On the other hand, if $p_{xz} = 0.8$, they vote for x

if $p_{xy} < 0.72$. Given that x beats z but y loses against z in the second round, these figures mean that voters 1–3 vote sincerely for y only if they have mistaken beliefs about the winning chances of the various alternatives.

Arrow's (1963) treatment assumes that all voters vote sincerely so that each one chooses the alternative that he or she prefers the most. Let $C_i(S)$ denote individual i's choice from a set of alternatives S and $\succ_i$ his or her preference ordering. Arrow (1963: 15) requires that the individual choices fulfil equation (7):[18]

$$C_i(S) = \{x|x \in S : \forall y \in S : x \succ_i y\}. \tag{7}$$

This condition is implicitly or explicitly present in all social-choice exercises that deal with preference aggregation. It requires that people vote sincerely.

Blin and Satterthwaite (1978) proved that if a voting procedure satisfies rationality (R), IIA and positive association (PA), then it also satisfies strategy-proofness. Given that Muller and Satterthwaite (1977) showed that strong positive association is equivalent to strategy-proofness, rationality and IIA together are sufficient conditions. As the amendment agendas are clearly not strategy-proof, this raises the question of whether it is rationality or IIA that is violated in the example: z was chosen when all voters voted sincerely, but x was chosen if some voted strategically. Two outcomes emerged from the single preference profile that were different from the two different behavioural assumptions. Does this mean that IIA is violated in the example? No, it does not.[19] To see this, let us have a closer look at Blin and Satterthwaite's framework.

Blin and Satterthwaite (1978) define a *voting procedure* (VP) as a function $v(\mathbf{p}|X)$ whose arguments are the profile of *stated* preferences $\mathbf{p}$ and the feasible set X. It is a single-valued mapping that selects one element of the feasible set to be the group's choice. They then define R, IIA and PA on voting procedures. A *social welfare function* (SWF) is any function u that gives, for any preference profile, a unique strict group preference ordering $P_N = u(\mathbf{p})$. A SWF $u(\mathbf{p})$ *underlies* a VP if and only if, for all profiles and all X, $v(\mathbf{p}|X) = \max_X[u(\mathbf{p})]$. A voting procedure satisfies *rationality* R if and only if there is an underlying SWF. A voting procedure satisfies IIA if and only if, for every feasible set X, $v(\mathbf{p}|X) = v(\mathbf{p}'|X)$ for all pairs of profiles $\mathbf{p}$ and $\mathbf{p}'$ for which all $x, y \in X$ and all $i \in N, x\mathbf{p}_i y$ if and only if $x\mathbf{p}'_i y$.

IIA is not violated in the example, even though there are two different outcomes from a single profile of *real* preferences because it only concerns how the *expressed* preferences are aggregated. IIA has nothing to do with how people behave when they vote. It merely says something about how the votes are computed to yield a social choice. In the example there really are *two* different profiles of expressed preferences, the sincere and the strategic, and because the ranking of alternatives x and z is not identical in these profiles, IIA is not violated because it does not apply.

IIA is satisfied even when there is strategic voting because amendment agendas compute the winner by making pairwise majority comparisons at each stage. As Saari (2008: 61) notes, the strategic voting argument is flawed simply because IIA does not preclude such voting.

Those who have presented the strategic voting argument might not give in so easily. What they might have in mind is something like the following intuitive argument for why IIA precludes strategic voting. If $C(\succ, S)$ and $C(\succ', S)$ in the definition of IIA above refer to the choices made under some voting rule, and if the preference profiles $\mathbf{p}$ and $\mathbf{p}'$ refer to real rather than expressed preferences, then the outcome determined in voting depends only on individual preferences for pairs of alternatives, and by implication strategic voting could not affect the choice under a voting rule that satisfies IIA (see, e.g., Vickrey, 1960). In fact, Blin and Satterthwaite recognize this viewpoint. They write as follows:

> In situations where incentives to engage in such manipulation do generally exist, then the design of acceptable voting procedures becomes difficult because a VP that gives acceptable choices when individuals honestly report their preferences may, relative to the individuals' true preferences, give unacceptable choices when individuals strategically misrepresent their preferences. Consequently, when we set up requirements , . . . , that a VP has properties such as R, IIA, PA, then we are assuming that individuals will in fact honestly report their true preferences. If we do not make this assumption, then we must construct a theory as to how individuals misrepresent their preferences. Suppose, given a particular VP $v(\mathbf{p}'|X)$, such a theory takes the form that $\mathbf{p}' = \omega(\mathbf{p}|X)$ where $\mathbf{p}'$ is the preference profile the individuals actually report for insertion into the voting procedure, $\mathbf{p}$ is the individuals' true preference profile , . . . , and ω is a function that describes how individuals misrepresent their true preferences. Given the functions v and ω, the function we really want to evaluate for acceptability is the composition of v and ω: $f(\mathbf{p}|X) = v[\omega(\mathbf{p}|X)|X]$. (Blin and Satterthwaite, 1978: 256)

In other words, what we are really interested in are the properties of voting procedures that take behavioural assumptions into account. Blin and Satterthwaite (1978: 257) also express what I presume to be a major methodological motivation for strategy-proofness. They say that one should check that a voting procedure satisfies strategy-proofness before any other properties are examined. The reason is that if we cannot guarantee strategy-proofness, we cannot guarantee whatever other properties we want our voting rules of the form $f(\mathbf{p}|X) = v[\omega(\mathbf{p}|X)|X]$ to satisfy.

In the example, a third 'irrelevant' alternative z affects the choice between x and y, intensities as well as beliefs matter for the result, and voters do not express their preferences sincerely. IIA is satisfied but it really does not imply *any* of the things people thought it did unless (7) and thereby strategy-proofness is also satisfied. The falsity of the strategic voting argument implies that even though a social choice function (SCF) satisfies IIA, the voting rule that is allegedly represented by the SCF does not even have the properties that the observability and the epistemological-moral argument attribute to the SCF. To put it differently, IIA might be argued for on the grounds of the observability or the epistemological-moral argument, or irrelevance or whatnot, if (7) were not violated. However, it is virtually always violated because strategy-proofness is virtually always violated. Thus, real voting procedures virtually never display the nice properties that IIA was supposed to provide. The only thing that IIA really requires is that once the ballots are cast, only pairwise information from those ballots should be taken into account. It follows that social choice functions which take into consideration the whole preference orderings, such as the (broad) Borda count, do not satisfy IIA.

Social choice theory is often distinguished from other approaches on the basis that condition (7) is satisfied. If it is satisfied, it makes sense to say that some social-choice functions satisfy some properties and some others do not. The fact that the Borda count violates IIA, whereas majority rule with agendas satisfies it matters very little when the corresponding voting procedures are evaluated normatively. IIA does not imply strategy-proofness or observability, and it does not even exclude intensity information in the voting rule unless strategy-proofness holds.

Saari (1998, 2001, 2003a, 2003b) argues that because IIA restricts the relevant information to preferences for pairs of alternatives, a voting rule that satisfies it does not allow for taking into account connecting information between the different pairs. Thus, even though one condition for Arrow's theorem explicitly requires voters to have transitive preferences, IIA implies that this transitivity information cannot be used. Hence, voting rules that satisfy IIA are incapable of distinguishing the cyclic preferences of (non-existent) irrational individuals from cyclic preference profiles. Saari's argument thus amounts to the idea that IIA is not normatively acceptable because voting rules that satisfy this condition fail to respect the rationality of voters.

However, if voters act strategically, they take the connecting information between pairs of alternatives into account in their behaviour.[20] The consequences of such behaviour are then reflected in the final outcomes. Even voting rules that satisfy IIA thus take connecting transitivity information between different pairs into account, albeit imperfectly and indirectly through strategic voting.

If the individual rationality argument against IIA is to be used for evaluating different voting rules, rules that do not satisfy IIA should be somehow better in providing full or direct information on such connecting rationality information. Saari proposes the *binary intensity IIA* as an alternative to IIA, and as a way of ensuring that rationality information is taken into account, and argues that the Borda count satisfies this condition (see also Saari, 1995). It requires that the aggregate ranking of each pair of alternatives is to be determined by each voter's relative ranking of that pair, and by the intensity of this ranking. The latter is determined by how many other alternatives are ranked between them. The problem with this condition is not just that, despite its name, it is best characterised as an ordinal one (Dowding, 2006; Risse, 2001), but also that the Borda count does not provide reliable intensity information on the rankings unless voters are sincere. The reader may verify that the Borda score is 7 for each of the three alternatives in the example if voters are sincere. However, if some voters act strategically, the Borda scores will be different. Suppose, for example, that voters 5 and 6 vote strategically by giving 2 points to x, 1 to z and 0 to y. Then the Borda scores are 9, 7 and 5 to x, y and z, respectively. This result provides false information on the intensity of the rankings. If the binary intensity IIA is also defined on the expressed votes, it is satisfied by the Borda count, but this does not mean that it will provide reliable information on the intensity of the rankings.

Consider, as another example, monotonicity under the Borda count. Suppose that the preferences are first given by Table 3 and all vote sincerely. The result will be a tie between all three alternatives. Then suppose that the popularity of z increases and that of x decreases because voter 7 changes his or her preferences from $x \succ_7 z \succ_7 y$ to $z \succ_7 x \succ_7 y$. Under sincere voting this change will make z the winner with scores 8, 7 and 6 for z, y and x, respectively. However, note that voters 1, 2 and 3 now have more reason to fear that their worst outcome z emerges unless they vote strategically for x. If they report the strategic orderings $x \succ y \succ z$, instead of the sincere $y \succ x \succ z$, x is selected with Borda score 9 even though it is less popular than before. Again, if monotonicity is defined on the ballots that have been cast or if the Borda count is defined as a social choice function that takes individual preferences as arguments, it satisfies monotonicity, but there is a very plausible scenario in which the outcome specified by the social choice function and the voting rule are not the same due to voters' strategic behaviour.

Satisfying IIA or its cognates is a non-issue, just like monotonicity (see also Austen-Smith and Banks (1991)). The general problem is that whether or not a social choice function satisfies this or that condition matters very little when strategy-proofness does not hold because the social-choice (or welfare) function on which the condition is defined does not adequately represent the corresponding voting procedure. This problem puts the very idea of social choice theory into question because the theory proceeds by imposing normative conditions on social choice functions.

If IIA does not preclude intensities and strategic voting, which of Arrow's conditions does? Two responses seem possible here. Assumption (7) could be taken to carry all the burden, or alternatively, one could say, as some scholars have done with respect to interpersonal comparability, that there is no single assumption or condition that rules them out. It is rather that the framework is set up in such a way that they are, in fact, never taken into account. Paraphrasing Sen (1970), it is the very idea that social choice theory uses functions (social choice, social welfare, for example) that take individual orderings as arguments that is responsible for precluding preference intensities.

Some condition or assumption in Arrow's theorem must be violated in these examples. The standard answer would be that it is the rationality of the social ordering because amendment agendas are pairwise voting procedures that satisfy IIA. This answer is only partly correct, however. Although there is a preference cycle in example 2, it cannot be a necessary reason for strategic voting because the example 1 does not have such a cycle. The crucial condition is rather (7). Rationality of the social ordering is violated merely because individuals are not rational in the sense of (7).

A defender of IIA might still wish to object and say that if $C(\succ, S)$ in the definition refers to actual choices in a voting procedure, and $\mathbf{p}$ and $\mathbf{p}'$ refer to real rather than expressed preferences, then IIA must prohibit strategic voting because it is violated in the examples. However, the examples show that $C(\succ, S)$ did not refer to the choice made under the voting rule, because (7) was violated. One might thus argue that there is no good reason to require that (7) must be satisfied for IIA to be applicable. After all, as the examples show, IIA and (7) are separate requirements because IIA does not imply strategy-proofness. IIA would thus be defined in terms of a voting rule of the form $f(\mathbf{p}|X) = v[\omega(\mathbf{p}|X)|X]$.

Independence of Irrelevant Alternatives for Voting Rules: If for all $x, y \in S$ and all individuals,

$$\mathbf{p}_{|S} = \mathbf{p}'_{|S} \rightarrow v[\omega(\mathbf{p}|X)|X] = v[\omega'(\mathbf{p}|X)|X], \tag{8}$$

where ω and ω' represent two different behavioural assumptions. In the example ω corresponds to sincere behaviour and ω' to expected utility maximisation. I take it that this definition is not acceptable as a definition of IIA because it includes voting rules with behavioural assumptions, but given the way in which IIA is now used in social choice theory, it only concerns the way in which preferences are aggregated. This is achieved either by considering functions that take preferences as arguments, or as in the case of Blin and Satterthwaite (1977, 1978), by using a function that denotes a voting procedure and by defining IIA in terms of the expressed ballots.

Suppose, however, for the sake of argument, that this were an acceptable definition of IIA. One might then argue that IIA is normatively justified if strategy-proofness is *also* assumed. If the only way to get the nice properties of IIA is by combining it with strategy-proofness, then let us always consider the two together.

This would be even worse for the normative acceptability of IIA. It would mean that IIA is violated in the examples due to strategic voting. However, the point of the examples is to illustrate that strategy-proofness is not desirable either. They show that strategic voting may have beneficial consequences if the utilitarian welfare function is used to evaluate the alternatives. The examples concern a voting rule that satisfies IIA, but a similar result is obtained in the Borda count, which does not satisfy IIA (Lehtinen, 2007a). Violations of strategy-proofness are thus normatively acceptable whether or not the underlying aggregation rule satisfies IIA.

8. Conclusion

Those who have not been willing to abandon IIA have emphasized that it is closely related to excluding strategic voting. However, from the utilitarian and thus welfarist point of view, strategic voting is desirable rather than undesirable under most commonly used voting rules. The reason for this is that it reflects preference intensities, and sincere voting does not allow for this under most voting rules. The strategy-proofness condition is thus not normatively acceptable under a utilitarian-welfarist evaluation. This is why strategic voting is not a convincing argument for IIA. Strategy-proofness has traditionally been defended on the basis of non-welfarist arguments. For example, it has been argued that unequal manipulative skills may lead to the destruction of efforts to design rules for the equal treatment of individuals.[21] The real trade-off is thus not between susceptibility to strategic manipulation, and rationality and intensities, but rather between welfarist arguments for and non-welfarist arguments against strategic voting.

Arrow's impossibility result and the closely related theorems given by Gibbard (1973) and Satterthwaite (1975) are unassailable as deductive proofs. However, we should not be concerned about these results because their most crucial conditions are not justifiable. Fortunately, we know that strategy-proofness is usually violated under all voting rules and that IIA does not preclude strategic voting.

Arrow's theorem also depends on condition (7). To the best of my knowledge, this condition has never been criticized in the literature on social choice, even though Arrow (1963) hinted that expected utility behaviour might solve the paradox he presented. However, if strategy-proofness is not justifiable, by implication (7) is not justifiable either. It is thus misleading to interpret Arrow's theorem as implying that there is something wrong with all voting procedures.

None of the arguments discussed in this paper (the epistemological-moral, the observability and the strategic-voting argument) is successful *as* an argument for using only preference orderings and Condorcet winners in voting theory. Restricting attention to Condorcet winners has always been justified not in genuine ethical arguments but by appealing to the observability or the epistemological-moral argument. There seems to be no good reason for evaluating voting outcomes on the basis of Condorcet winners rather than utilitarian winners if such pragmatic arguments fail. Utilitarian winners are to be preferred on genuine ethical grounds, however, because they take preference intensities into account.

The above reasoning thus gives rise to four methodological conclusions. First, given that the three main arguments for IIA and against intensities fail, there is no reason to favour Condorcet winners over utilitarian winners in welfarist evaluations of voting rules. Second, the notion of cyclic preferences and the absence of a Condorcet winner have been given an all-too-prevalent role in voting theory. The possibility that preferences are cyclic is only one among many factors that may influence voting outcomes. Beliefs, information and preference intensities are also important. Models that take into account only preference orderings provide a misleading picture of voting rules in that they are based on the false empirical assumption that voting is characterized by choice under certainty.

Third, the very idea of imposing conditions on social choice (or welfare) functions under the assumption that voters act sincerely is suspect because we can never be sure that actual voting rules will have the properties that the conditions were supposed to guarantee. Thus, showing that a social choice function satisfies this or that set of conditions is irrelevant for a comparison between different social choice functions. Note also that since the counterbalancing model fulfills the conditions for an invisible-hand model (Lehtinen, 2009), the criticism of strategy-proofness could not have been derived with the methods of social choice theory because the very notion of aggregativity rules out such models.

Finally, the theory of strategic voting has not addressed the right questions. If strategic voting is beneficial under many commonly used voting rules, it is not very fruitful to seek strategy-proof voting mechanisms or to find out which voting rules are least susceptible to it. The relevant question concerns how much strategic voting increases (or perhaps decreases) utilitarian efficiency under various voting rules under different assumptions concerning voters' willingness to take risks and preference intensities. There are significant differences between different voting rules in these respects.

Acknowledgements

I am grateful to Hannu Nurmi, Eerik Lagerspetz and Martin van Hees for their comments on previous versions of this paper. The usual disclaimer applies. This paper is a substantially revised version of the concluding chapter ('A farewell to IIA') of my doctoral dissertation (The welfare consequences of strategic voting, EIPE, Erasmus University Rotterdam, 2007). It has been presented at the European Philosophy of Science Association conference in Madrid 2007, the Tilburg Social Choice Colloquium in 2007, the TINT Seminar in Helsinki 2008 and at the European Public Choice Society Conference in Izmir 2010.

Funding

The research was funded by the Academy of Finland.

Notes

1. See, e.g., McKelvey and Ordeshook (1972), Enelow (1981), Myerson and Weber (1993), Cox (1997) and Myatt (2007).
2. See also Rawls (1971) and Pattanaik (1968).
3. He explicitly excluded strategic considerations (Arrow, 1963).

4. For example, Satterthwaite (1975) shows that strategy-proof voting procedures are equivalent to social-welfare functions that satisfy citizen sovereignty, non-negative response and IIA.
5. For recent papers presenting this argument see, e.g., McLean (1987). Arrow (1977) also puts it forward. Saari first (2001, 2008) acknowledges and then refutes. Even some of those who do not espouse IIA think that strategy-proofness follows from it (Mackie, 2003).
6. A choice function assigns a choice in each environment S; $C(S) = \{x \mid \text{ all } y \text{ in } S : xRy\}$.
7. See, e.g., Fishburn (1989).
8. Here I am disregarding the entirely different question of whether the riskiness of the *choice alternatives* in an election should be taken into account.
9. See also Arrow (1963).
10. If indifference is ruled out by assuming that $0 < v^i < 1$ for all voters, this intensity must also be cardinally stronger.
11. This expression can be derived by setting $V(\text{Max}) = a + bU(\text{Max}) = 1$, $V(\text{Min}) = 0$ and $V(\text{Med}) = v$, and solving for a and b.
12. Sen (1976, 1977, 1986), Weymark (1991, 2005) and Roemer (1996) are the critical protagonists. Broome (1991, 2008b), Gibbard (2008), Ng (1999) and Risse (2002) could be counted as defenders of Harsanyi's position. See also Mongin (2001).
13. I cannot ignore the failure of Harsanyi's argument altogether, however, because I need to explain what gives us the right to use the *sum* of utilities as the welfare function rather than some other functional form such as the product. The reason is that using the product would make the results of the models depend on the morally arbitrary fact concerning whether or not there is an individual in the population whose utility is exactly or very close to zero. In such cases the product would be zero, or it would depend too much on one single person's utility.
14. See Sen (1970), Mackay (1980: 79) and Kemp and Ng (1987).
15. If IIA is formulated in such a way that it refers to cardinal-utility profiles, we end up with an impossibility result because cardinal utility without interpersonal comparisons does not make the impossibility result vanish (Kalai and Schmeidler, 1977; Sen, 1970). Accordingly, the standard view is that the most reasonable way to eschew Arrovian impossibility is to make interpersonal comparisons.
16. Mackie (2003) provides a detailed overview of such criticisms.
17. The utilities in this example are identical to those that Arrow (1963: 32) used to criticize 'utilitarian' voting and argue for IIA. Only the labelling is different: Arrow's example is obtained by interchanging x and y.
18. The assumption that all preferences are strict is used here. Given that Arrow also requires that the social ordering is rational, he does not put indices indicating an individual into the equation.
19. Depending on how exactly IIA is defined, it might be violated under certain agendas. Mbih and Moyouwou (2008) allow for changes in the number of voters, and this version of IIA is violated under amendment agendas.
20. I emphasized this issue to Saari in a discussion in a conference in 2006, but for some reason, he did not mention this discussion when he makes this point in Saari (2008: 60).
21. See Kelly (1988). These arguments were originally presented in Mark Satterthwaite's PhD dissertation: The Existence of a Strategy Proof Voting Procedure (University of Wisconsin, 1973). See van Hees and Dowding (2008) for a comprehensive discussion.

References

Arrow KJ (1951) *Social Choice and Individual Values*. New York: Wiley–Chapman & Hall.

Arrow KJ (1963) *Social Choice and Individual Values*, 2nd ed. New Haven, CT: Yale University Press.

Arrow KJ (1973a) Formal theories of social welfare. In: Wiener P (ed.) *Dictionary of the History of Ideas*. New York: Charles Scribner's Sons.
Arrow KJ (1973b) Some ordinalist-utilitarian notes on Rawls' Theory of Justice. *Journal of Philosophy* 70: 245–263.
Arrow KJ (1977) Current developments in theory of social choice. *Social Research* 44(4): 607–622.
Arrow KJ (1983 [1967]) *Collected Papers of Kenneth Arrow, Vol 1; Social Choice and Justice*. Cambridge, MA: Belknap Press.
Arrow KJ (1987) Reflections on the essays. In: Feiwel GR (ed.) *Arrow and the Foundations of the Theory of Economic Policy*. UK: Macmillan.
Arrow KJ (1997) The functions of social choice theory. In: Arrow KJ, Sen AS and Suzumura K (eds) *Social Choice Re-Examined*. London: Macmillan.
Austen-Smith D and Banks J (1991) Monotonicity in electoral systems. *American Political Science Review* 85(2): 531–537.
Blin J-M and Satterthwaite MA (1977) On preferences, beliefs, and manipulation within voting situations. *Econometrica* 45(4): 881–888.
Blin J-M and Satterthwaite MA (1978) Individual decisions and group decisions. *Journal of Public Economics* 10(2): 247–268.
Bordes G and Tideman NT (1991) Independence of irrelevant alternatives in the theory of voting. *Theory and Decision* 30(2): 163–186.
Broome J (1991) *Weighing Goods: Equality, Uncertainty and Time*. Cambridge, MA: Basil Blackwell.
Broome J (2008a) Can there be a preference-based utilitarianism?' In: Fleurbaey M, Salles M and Weymark JA (eds) *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*. Cambridge: Cambridge University Press.
Broome J (2008b) Comments on Allan Gibbard's Tanner Lectures. In: Stroud B (ed.) *Reconciling Our Aims*. Oxford: Oxford University Press.
Coleman JS (1966) The possibility of a social welfare function. *American Economic Review* 56(5): 1105–1022.
Cox GW (1997) *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. New York: Cambridge University Press.
Dowding KM (2006) Can populism be defended? William Riker, Gerry Mackie and the interpretation of democracy. *Government and Opposition* 41(3): 327–346.
Enelow JM (1981) Saving amendments, killer amendments, and an expected utility theory of sophisticated voting. *Journal of Politics* 43(4): 1062–1089.
Fishburn PC (1989) Retrospective on the utility theory of Von Neumann and Morgenstern. *Journal of Risk and Uncertainty* 2(2): 127–158.
Gibbard AF (1973) Manipulation of voting schemes: a general result. *Econometrica* 41(4): 587–601.
Gibbard AF (2008) *Reconciling Our Aims: In Search of Bases for Ethics*. Oxford: Oxford University Press.
Hammond PJ (1987) On reconciling Arrow's Theory of Social Choice with Harsanyi's Fundamental Utilitarianism. In: Feiwel GR (ed.) *Arrow and the Foundations of the Theory of Economic Policy*. London: Macmillan.
Hansson B (1973) The independence condition in the theory of social choice. *Theory and Decision* 4: 25–49.
Harsanyi JC (1953) Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* 61: 434–435.
Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63(4): 309–321.

Harsanyi JC (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. New York: Cambridge University Press.

Harsanyi JC (1978) Bayesian decision-theory and utilitarian ethics. *American Economic Review* 68(2): 223–228.

Harsanyi JC (1979) Bayesian decision-theory, rule utilitarianism, and Arrow impossibility theorem. *Theory and Decision* 11(3): 289–317.

Harsanyi JC (1987) Von Neumann–Morgenstern utilities, risk taking, and welfare. In: Feiwel GR (ed.) *Arrow and the Ascent of Modern Economic Theory*. London: Macmillan.

Harsanyi JC (1992) Game and decision theoretic models in ethics. In: Aumann RJ and Hart S (eds) *Handbook of Game Theory with Economic Applications*. New York: Elsevier.

Harsanyi JC (1993) Normative validity and meaning of Von Neumann–Morgenstern utilities. In Binmore K, Kirman AP and Tani P (eds) *Frontiers of Game Theory*. Cambridge, MA: MIT Press.

Hillinger C (2005) The case for utilitarian voting. *Homo Economicus* 22(3): 295–321.

Hirshleifer J and Riley JG (1992) *The Analytics of Uncertainty and Information*. New York: Cambridge University Press.

Kalai E and Schmeidler D (1977) Aggregation procedure for cardinal preferences: a formulation and proof of Samuelson's Impossibility Conjecture. *Econometrica* 45(6): 1431–1438.

Kelly JS (1988) *Social Choice Theory: An Introduction*. New York: Springer-Verlag.

Kemp MC and Ng Y-K (1987) Arrow's independence condition and the Bergson–Samuelson tradition. In Feiwel GR (ed.) *Arrow and the Foundations of the Theory of Economic Policy*. New York: New York University Press.

Lehtinen A (2006) Signal extraction for simulated games with a large number of players. *Computational Statistics and Data Analysis* 50: 2495–2507.

Lehtinen A (2007a) The Borda rule is also intended for dishonest men. *Public Choice* 133(1–2): 73–90.

Lehtinen A (2007b) The welfare consequences of strategic voting in two commonly used parliamentary agendas. *Theory and Decision* 63(1): 1–40.

Lehtinen A (2008) The welfare consequences of strategic behaviour under approval and plurality voting. *European Journal of Political Economy* 24: 688–704.

Lehtinen A (2009) Intentions in invisible-hand accounts. *Journal of Economic Methodology* 16(4): 409–416.

Lehtinen A (2010) Behavioural heterogeneity under approval and plurality voting. In: Laslier J-F and Sanver MR (eds) *Handbook on Approval Voting*. Heidelberg: Springer.

MacKay AF (1980) *Arrow's Theorem, the Paradox of Social Choice: A Case Study in the Philosophy of Economics*. New Haven, CT: Yale University Press.

Mackie G (2003) *Democracy Defended*. Cambridge: Cambridge University Press.

Mbih B and Moyouwou I (2008) Violations of independence under amendment and plurality rules with anonymous voters. *Group Decision and Negotiation* 17(4): 287–302.

McKelvey RD and Ordeshook PC (1972) A general theory of the calculus of voting. In: Herndon JF and Bernd JL (eds) *Mathematical Applications in Political Science*. Charlottesville, VA: University Press of Virginia.

McLean I (1987) *Public Choice: An Introduction*. Oxford: Basil Blackwell.

Mongin P (2001) The impartial observer theorem of social ethics. *Economics and Philosophy* 17(2): 147–179.

Muller E and Satterthwaite MA (1977) The equivalence of strong positive association and strategy-proofness. *Journal of Economic Theory* 14(2): 412–418.

Munger M (2009) Donald G. Saari, disposing dictators, demystifying voting paradoxes: social choice analysis. *Public Choice* 140(3): 539–542.

Myatt DP (2007) On the theory of strategic voting. *Review of Economic Studies* 74(1): 255–281.

Myerson RB and Weber RJ (1993) A theory of voting equilibria. *American Political Science Review* 87(1): 102–114.
Ng Y-K (1999) Utility, informed preference, or happiness: following Harsanyi's argument to its logical conclusion. *Social Choice and Welfare* 16(2): 197–216.
Pattanaik PK (1968) Risk, impersonality, and the social welfare function. *Journal of Political Economy* 76(6): 1152–1169.
Plott CR (1976) Axiomatic social choice theory: an overview and interpretation. *American Journal of Political Science* 20(3): 511–596.
Rawls J (1971) *A Theory of Justice*. Cambridge, MA: Belknap Press.
Risse ME (2001) Arrow's theorem, indeterminacy, and multiplicity reconsidered. *Ethics* 111(4): 706–734.
Risse ME (2002) Harsanyi's 'Utilitarian Theorem' and utilitarianism. *Nous* 36(4): 550–577.
Roemer JE (1996) *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
Roemer JE (2008) Harsanyi's impartial observer is not a utilitarian. In: Fleurbaey M, Salles M and Weymark JA (eds) *Justice, Political Liberalism, and Utilitarianism; Themes from Harsanyi and Rawls*. Cambridge: Cambridge University Press.
Saari DG (1995) *Basic Geometry of Voting*. Berlin: Springer.
Saari DG (1998) Connecting and resolving Sen's and Arrow's theorems. *Social Choice and Welfare* 15(2): 239–261.
Saari DG (2001) *Decisions and Elections: Explaining the Unexpected*. Cambridge: Cambridge University Press.
Saari DG (2003a) Capturing the 'Will of the People'. *Ethics* 113(2): 333–349.
Saari DG (2003b) Unsettling aspects of voting theory. *Economic Theory* 22(3): 529–555.
Saari DG (2008) *Disposing Dictators, Demystifying Voting Paradoxes: Social Choice Analysis*. Cambridge: Cambridge University Press.
Satterthwaite MA (1975) Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10(2): 187–217.
Sen AK (1970) *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
Sen AK (1976) Welfare inequalities and Rawlsian axiomatics. *Theory and Decision* 7(4): 243–262.
Sen AK (1977) Non-linear social welfare functions: a reply to Professor Harsanyi. In: Butts RE and Hintikka J (eds) *Foundational Problems in the Special Sciences*. Dordrecht: Reidel.
Sen AK (1986) Social choice theory. In: Arrow KJ and Intriligator M (eds) *Handbook of Mathematical Economics*. Heidelberg: Elsevier.
Smith W (2000) Range voting. http://www.math.temple.edu/˜wds/homepage/rangevote.pdf.
Strasnick S (1976) The problem of social choice: Arrow to Rawls. *Philosophy & Public Affairs* 5(3): 241–273.
van Hees M and Dowding KM (2008) In praise of manipulation. *British Journal of Political Science* 38: 1–15.
Vickrey W (1960) Utility, strategy, and social decision rules. *Quarterly Journal of Economics* 74(4): 507–535.
Weymark JA (1991) A reconsideration of the Harsanyi–Sen debate on utilitarianism. In: Elster J and Roemer JE (eds) *Interpersonal Comparisons of Well-being*. Cambridge: Cambridge University Press.
Weymark JA (2005) Measurement theory and the foundations of utilitarianism. *Social Choice and Welfare* 25(2–3): 527–555.