4
Interval-Scale Theories and Variance Voting

Introduction

In Chapter 3, we discussed how to take into account moral uncertainty over merely ordinal and non-comparable theories. But, very often, theories will provide interval-scale measurable choice-worthiness functions. This chapter discusses how to take into account moral uncertainty over interval-scale measurable but non-comparable theories. Once again, we make use of the analogy between decision-making under moral uncertainty and voting.

In section I, we give examples of interval-scale theories where it's plausible to think that these theories are incomparable with each other. From section II onwards, we discuss what to do in such cases. In section II, we consider but reject the idea that one should use the Borda Rule in such situations. We then consider Ted Lockhart's idea that, in conditions of intertheoretic incomparability, one should treat each theory's maximum and minimum degree of choice-worthiness within a decision-situation as equal, and then aggregate using MEC. This is the analogue of *range voting*.

We consider Sepielli's objection that the principle is arbitrary, but argue that the idea of giving every theory 'equal say' has the potential to make the account non-arbitrary. However, in section III, we argue that Lockhart's suggestion fails by this principle, and that what we call *variance voting* is uniquely privileged as the account that gives incomparable theories equal say. We give intuitive examples in favour of this view, and then show, in section IV, that on either of two ways of making the principle of 'equal say' precise it is only variance voting that gives each theory 'equal say'.

In section V, we discuss what to do in conditions where one has positive credence in some merely ordinal theories, some interval-scale but non-comparable theories, and some theories that are both interval-scale measurable and comparable with each other. In section VI, we discuss whether

Moral Uncertainty. William MacAskill, Krister Bykvist and Toby Ord, Oxford University Press (2020).

DOI: 10.1093/oso/9780198722274.001.0001

the normalization used by this account should be done only within the decision-situation at hand, or whether it should be done over all possible decision-situations.

I. Intertheoretic Incomparability

As described in Chapter 3, a problem that has dogged accounts of decision-making under moral uncertainty is how to make intertheoretic comparisons of choice-worthiness differences.[1] Describing this more fully, the problem is as follows. All a moral theory needs to provide, one might suppose, is all the true statements of the form, '*A* is at least as choiceworthy as *B*', where *A* and *B* represent possible options.[2]

If the choice-worthiness relation of the moral theory orders all options (including lotteries) and satisfies the von Neumann–Morgenstern axioms, then we can construct an interval-scale measure of choice-worthiness.[3] This means that we can represent this choice-worthiness relation using a choice-worthiness function so that it's meaningful to say that the difference in choice-worthiness between two options *A* and *B*, according to the theory, is greater than, less than, or equal to, the difference in choice-worthiness between two other options *C* and *D*.[4] But, importantly, the choice-worthiness function is only unique up to a positive affine transformation: if you multiply that numerical representation by a positive constant or add any constant, the resulting function still represents the same choice-worthiness ordering. Thus, from the moral theories alone, even though we can meaningfully compare differences of choice-worthiness *within* a moral theory, we just don't have enough information to enable us to compare differences of choice-worthiness *across* moral theories.[5] But if so, then we cannot apply MEC.

[1] The problem is normally called the 'problem of intertheoretic comparisons of value'. But this is somewhat misleading. What we and the others who have explored decision-making under moral uncertainty are primarily interested in is comparing *choice-worthiness* across moral theories, rather than comparing *value* across theories.

[2] We deny this supposition in the following chapter; but assuming it provides a particularly clear way of understanding of where the problem comes from.

[3] Namely: Transitivity, Completeness, Continuity, and Independence. For discussion of these axioms in relation to moral theory, see Broome, *Weighing Goods*.

[4] In fact, it even allows us to talk about the ratio between two such differences.

[5] A similar problem arises in the study of social welfare in economics: it is desirable to be able to compare the strength of preferences of different people, but even if you represent

There are really (at least) two questions that fall under the label of 'the problem of intertheoretic choice-worthiness comparisons'. The first question is:

When, if ever, are intertheoretic choice-worthiness comparisons (of differences) possible, and in virtue of what are true intertheoretic comparisons true?

We address this question in Chapter 5. The second question is:

Given that choice-worthiness (of differences) is sometimes incomparable across first-order moral theories, what is it appropriate to do in conditions of moral uncertainty?

We focus on this second question in this chapter, addressing the situation where the non-comparable theories are interval-scale measurable.

To show that it's at least plausible that theories are sometimes interval-scale measurable but incomparable, let's consider two consequentialist theories, prioritarianism and utilitarianism. Prioritarianism gives more weight to gains in wellbeing to the worse-off than it does to gains in wellbeing to the better-off. But does it give more weight to gains in wellbeing to the worse-off than utilitarianism does? That is, is prioritarianism like utilitarianism but with additional concern for the worse-off; or is prioritarianism like utilitarianism but with *less* concern for the better-off? We could represent the prioritarian's idea of favouring the worse-off over the better-off equally well either way. And there seems, at least, to be no information that could let us determine which of these two ideas is the 'correct' way to represent prioritarianism vis-à-vis utilitarianism.

Now, one might think that there is an easy solution, relying on the fact that both of these views make the same recommendations in situations that involve saving identical lives under uncertainty. On both views, a 50% chance of saving two lives with the same lifetime wellbeing and a guarantee of saving one of those lives are equally choiceworthy. So, according to both of these theories, saving two identical lives is twice as good as saving one. One might think that one can use this 'agreement' between the two theories on the

preferences by interval-scale measurable utility functions, you need more information to make them comparable.

difference in choice-worthiness between saving one life and saving two as a common measure.[6]

To see that this doesn't work, consider Annie and Betty. For each of these people, if you administer a certain drug they'll each live for nine more years. Both utilitarianism and prioritarianism agree that the difference in choice-worthiness between doing nothing and saving both Annie and Betty is exactly twice as great as the difference in choice-worthiness between doing nothing and saving Annie alone. For concreteness, we'll assume that the prioritarian's concave function is the square root function. And we'll begin by assuming that Annie and Betty have lived for sixteen years so far. If so, then the prioritarian claims that the choice-worthiness difference between saving both Betty and Annie's lives and saving Annie's life alone is $\sqrt{25}-\sqrt{16}$, which equals 1. The utilitarian claims that this difference is $25-16$, which equals 9. So if we are normalizing the two theories at the difference between saving one life and saving two, then 1 unit of choice-worthiness, on prioritarianism, equals 9 units of choice-worthiness, on utilitarianism.

But now suppose that both Annie and Betty had lived much longer. Suppose they had lived for sixty-four years each. In this case, the difference in choice-worthiness, on prioritarianism, between saving both Betty and Annie's lives, and saving Annie's life alone is $\sqrt{73}-\sqrt{64}$, which is approximately 0.5. The utilitarian, in contrast, claims that this difference is $73-64$, which equals 9. So, if we are normalizing the two theories at the difference between saving one life and saving two in this case, then 1 unit of choice-worthiness, on prioritarianism, equals approximately 18 units of choice-worthiness, on utilitarianism. But this is inconsistent with our previous conclusion. Applying the 'normalize at the difference between saving one life and saving two' rule gives different answers depending on which two lives we're talking about.

So we cannot consistently normalize utilitarianism and prioritarianism at the difference ratio between saving one life and saving two lives, and saving two lives and saving no lives. With this possibility ruled out, it thus seems very difficult to see how there could be any principled way of claiming that there is a unit of value that is shared between utilitarianism and prioritarianism. So one might reasonably think that they cannot be placed on a common scale.

[6] Both Ross, 'Rejecting Ethical Deflationism', p. 764 and Sepielli, 'What to Do When You Don't Know What to Do' make suggestions along these lines. We discuss this idea more thoroughly in the next chapter.

This gives at least one case where choice-worthiness differences seem, on their face, to be incomparable between different theories. But if we have no way of making the intertheoretic comparison, then we cannot take an expectation over those moral theories. Given this, it's unclear what a decision-maker under moral uncertainty should do if she faces theories that are interval-scale measurable but intertheoretically incomparable. So we need an account of what it's appropriate to do in conditions where we cannot put two different moral theories on a common scale. Let us now look at some contenders.

II. Two Unsatisfactory Proposals

One might initially think that our work in Chapter 3 gives a solution. When theories are intertheoretically incomparable, one should aggregate those theories' choice-worthiness orderings using the Borda Rule.

The problem with this proposal should be obvious. Consider the decision-situation in Table 4.1.

In this case, the difference between *B* and *C*, on T_1, is far greater than the difference between *A* and *B*. Similarly, the difference between *A* and *B*, on T_2, is far greater than the difference between *B* and *C*. Yet the difference between the Borda Scores of *A* and *B* is the same as the difference in the Borda Scores between *B* and *C*, on both theories. The Borda Rule therefore seems to misrepresent the theories themselves, throwing away interval-scale information when we have it. The voting analogy might prove useful, but ignoring interval-scale information when we have it is not the way to proceed.

Lockhart has suggested a different account: what he calls the 'Principle of Equity among Moral Theories'. He defines it as follows:

> The maximum degrees of moral rightness of all possible actions in a situation according to competing moral theories should be considered equal. The minimum degrees of moral rightness of possible actions in a situation according to competing theories should be considered equal unless all possible actions are equally right according to one of the theories (in which case all of the actions should be considered to be maximally right according to that theory).[7]

[7] Lockhart, *Moral Uncertainty and Its Consequences*, p. 84.

Table 4.1

	T_1—50%	T_2—50%
A	10	0
B	9	90
C	0	100

It's ambiguous whether Lockhart thinks that the PEMT is giving an account of how two theories actually compare, or whether he is giving an account of what to do, given that all theories are incomparable. In the above quote it sounds like the latter, because he says, 'should be considered' rather than 'is', and this is how we'll understand it in this chapter. (In Chapter 5 we will consider whether accounts similar to Lockhart's are plausible as accounts of how choice-worthiness actually compares intertheoretically, and argue that they are not.)

Lockhart's account is analogous to range voting.[8] On range voting, every voter can give each candidate a score, which is a real number from, say, 0 to 10. The elected candidate is the candidate whose sum total of scores across all voters is highest.

To illustrate Lockhart's account, let's look again at the previous table. If we were to take the numbers in the table at face value, then we would suppose that the difference between *B* and *C*, on T_2, is ten times as great as the difference between *A* and *B*, on T_1. But to do so would be to forget that each theory's choice-worthiness function is unique up to its own positive affine transformation. According to Lockhart's proposal, we should treat the best and worst options as equally choiceworthy. So we should treat the choice-worthiness of $CW_1(A)$ as the same as the choice-worthiness of $CW_2(C)$ and we should treat the choice-worthiness of $CW_1(C)$ as the same as the choice-worthiness of $CW_2(A)$ (using '$CW_n(A)$' to refer to the number assigned to option *A* by theory *n*'s choice-worthiness function). One way of representing the theories, therefore, in accordance with the PEMT is as in Table 4.2.

What seems promising about Lockhart's account is that it provides a way of taking into account moral uncertainty across interval-scale theories that are incomparable. However, Lockhart's account has come under fire

[8] Claude Hillinger, 'The Case for Utilitarian Voting', *Homo Oeconomicus*, vol. 22, no. 3 (2005), pp. 295–321.

Table 4.2

	T_1—50%	T_2—0.5%
A	10	0
B	9	9
C	0	10

in a recent article by Andrew Sepielli.[9] Most of the problems with his account arise from the fact that it treats maximum and minimum degrees of choice-worthiness as the same *within a decision-situation* rather than *across all possible decision-situations*. We discuss those criticisms in section VI; in the meantime, we'll stick with the *within a decision-situation* formulation.

For now, we want to discuss a different problem that Sepielli raises. As he puts it, 'perhaps the most telling problem with the PEMT is that it is arbitrary'.[10]

There is a wide array of alternatives to Lockhart's view. Why, one might ask, should one treat the maximum and minimum choiceworthiness as the same, rather than the difference between the most choiceworthy option and the mean option, or between the least choice worthy option and the mean option? Or why not treat the mean difference in choiceworthiness between options as the same for all theories?

Lockhart anticipates this objection, stating: 'It may appear that I have, in an ad hoc manner, concocted the PEMT for the sole purpose of defending the otherwise indefensible claim that moral hedging is possible'.[11] However, he responds as follows.

> The PEMT might be thought of as a principle of fair competition among moral theories, analogous to democratic principles that support the equal counting of the votes of all qualified voters in an election regardless of any actual differences in preference intensity among the voters…PEMT appears not to play favorites among moral theories or to give some type(s) of moral theories unfair advantages over others.[12]

[9] Andrew Sepielli, 'Moral Uncertainty and the Principle of Equity among Moral Theories', *Philosophy and Phenomenological Research* 86, no. 3 (2013), pp. 580–9.

[10] Sepielli, 'Moral Uncertainty and the Principle of Equity among Moral Theories', p. 587.

[11] Lockhart, *Moral Uncertainty and Its Consequences*, p. 86.

[12] Lockhart, *Moral Uncertainty and Its Consequences*, p. 86.

That is, he appeals to what we'll call the *principle of 'equal say'*: the idea, stated imprecisely for now, that we want to give equally likely incomparable moral theories equal weight when considering what it's appropriate to do, and that the degree of influence that a moral theory has over the appropriateness of options across a wide variety of different decision-situations should be only in proportion to the degree of credence assigned to that theory.

As Sepielli points out, this idea doesn't seem at all plausible if we're trying to use the PEMT as a way of actually making intertheoretic comparisons. Considerations of fairness are relevant to issues about how to treat *people*: one can be unfair to a person. But one cannot be unfair to a theory. Perhaps by saying that one was being 'unfair' to Kantianism, one could mean that one's degree of belief was too low in it. But one can't be unfair to it insofar as it 'loses out' in the calculation of what it's appropriate to do. If a theory thinks that a situation is low stakes, we should represent it as such.

But the idea of 'equal say' has more plausibility if we are talking about how to come to a decision in the face of genuine intertheoretic incomparability. In developing an account of decision-making under moral uncertainty, we want to remain neutral on what the correct moral theory is: we do not want to bias the outcome of the decision-making in favour of some theories over others. Against this one could argue that some theories are simply higher stakes in general than other theories. But if, as we assume in this chapter, we are in a condition where there really is no fact of the matter about how two theories compare, then we cannot make sense of the idea that things might be higher stakes in general for one theory rather than the other. So we need a way of taking uncertainty over those theories into account that is not biased towards one theory rather than another.

To see a specific case of how this could go awry, consider average and total utilitarianism, and assume that they are indeed incomparable. Suppose that, in order to take an expectation over those theories, we choose to treat them as agreeing on the choice-worthiness of differences between options in worlds where the only person that exists is the decision-maker, and therefore only their welfare is at stake. If we do this, then, for almost all practical decisions about population ethics, the appropriate action will be in line with what total utilitarianism regards as most choiceworthy because, for almost all decisions (which involve a world with billions of people), the stakes would be large for total utilitarianism, but tiny for average utilitarianism. So it is plausible that, if we treat the theories in this way, we are being partisan towards total utilitarianism.

In contrast, if we chose to treat the two theories as agreeing on the choice-worthiness differences between options with worlds involving some extremely large number of people (say 10^{100}), then for almost all real-world decisions, what it is appropriate to do will be the same as what average utilitarianism regards as most choiceworthy. This is because we are representing average utilitarianism as claiming that, for almost all decisions, the stakes are much higher than for total utilitarianism. In which case, it seems that we are being partisan to average utilitarianism. What we really want is to have a way of treating the theories such that each theory gets equal influence.

Lockhart states that the PEMT is the best way to give every theory 'equal say'. But he doesn't argue for that conclusion, as Sepielli notes:[13]

> But even granting that some 'equalization' of moral theories is appropriate, Lockhart's proposal seems arbitrary. Why equalize the maximum and minimum value, rather than, say, the mean value and the maximum value? […] It seems as though we could find other ways to treat theories equally, while still acknowledging that the moral significance of a situation can be different for different theories. Thus, even if we accept Lockhart's voting analogy, there is no particularly good reason for us to use PEMT rather than any of the other available methods.

In a very similar vein, Amartya Sen has argued against an analogue of the PEMT within social choice theory, the 'zero-one' rule:[14]

> It may be argued that some systems, e. g., assigning in each person's scale the value 0 to the worst alternative and the value 1 to his best alternative are interpersonally 'fair' but such an argument is dubious. First, there are other systems with comparable symmetry, e.g., the system we discussed earlier of assigning 0 to the worst alternative and the value 1 to the sum of utilities from all alternatives.

We think both Sen and Sepielli are right that principled reasons for endorsing the PEMT over its rivals have not been given. But, further to that, we will argue in the following two sections that it's demonstrably *false* that the PEMT is the best way of giving each theory 'equal say'. Instead, we think

[13] Sepielli, 'Normative Uncertainty for Non-Cognitivists', pp. 587–8.
[14] Sen, *Collective Choice and Social Welfare*, p. 98.

that what we'll call variance voting is the best way to take moral uncertainty into account across theories that are interval-scale and incomparable, because it is the best way of giving each theory 'equal say'.

III. Variance Voting

We'll call Lockhart's view and its rivals interval-scale *voting systems*. To develop an intuitive sense of how different interval-scale voting systems can differ in how they apportion 'say' between theories, let's consider some examples. Let's consider four different interval-scale voting systems using the 'across all decision-situations' formulation of each:

(i) Lockhart's PEMT, which treats the range of the choice-worthiness function (i.e. the difference between minimum and maximum assigned values) as the same across all interval-scale and incomparable theories;
(ii) what we'll call max-mean, which treats the difference between the mean choice-worthiness and the maximum choice-worthiness as the same across all interval-scale and incomparable theories;
(iii) what we'll call mean-min, which treats the difference between the mean choice-worthiness and the minimum choice-worthiness of all interval-scale and incomparable theories as the same (this is the account that Sen suggests in the above quote);
(iv) variance voting, which treats the variance (i.e. the average of the squared differences in choice-worthiness from the mean choice-worthiness) as the same across all theories.

Variance is a very important statistical property, measuring how spread out choice-worthiness is over different options. While its formula is a bit more complex, it is typically seen as the most natural measure of spread. Since the variance is the square of the standard deviation, normalizing at variance is the same as normalizing at the size of the standard deviation.[15] One can

[15] In order to make sense of the variance of a choice-worthiness function, we need a notion of *measure* over possibility space. This is discussed in section V, in relation to the Borda Rule. We assume that we should use the same choice of measure when using variance voting as we do when using the Borda Rule. Having a measure over the option-set allows variance normalization to apply to many unbounded moral theories: we take this to be yet another advantage of variance voting over the PEMT.

compute the normalized choice-worthiness for an option by subtracting the mean choice-worthiness, then dividing by the standard deviation.

Note that like Lockhart's original statement of PEMT, we should, for each of these normalization methods, also specify that if a theory ranks all options as exactly equally choiceworthy, all of these four methods leave its choice-worthiness function alone: the normalized choice-worthiness function is just equal to the original one. To do otherwise would involve dividing by zero.

We shall apply these four different structural normalization methods to four types of first-order moral theory. We'll call the first type *Bipolar* theories. According to Bipolar theories, the differences in choice-worthiness among the most choiceworthy options, and among the least choiceworthy options, are zero or tiny compared to the differences in choiceworthiness between the most choiceworthy options and the least choiceworthy options. For example, a view according to which violating rights is impermissible, everything else is permissible, and where there is very little difference in the severity of wrongness between different wrong actions, would be a Bipolar theory.

We'll call the second type of theory *outlier* theories. According to this view, most options are roughly similar in choiceworthiness, but there are some options that are extremely choiceworthy, and some options that are extremely un-choiceworthy. A bounded total utilitarian view with a very high and very low bounds might be like this: the differences in value between most options are about the same, but there are some possible worlds which, though unlikely, are very good indeed, and some other worlds which, though unlikely, are very bad indeed.

We'll call the third type of theory *Top-Heavy*. According to this type of theory, there are a small number of outliers in choice-worthiness, but they are only on one side of the spectrum: there are just a small number of extremely un-choiceworthy possible options. Any consequentialist theory that has a low upper bound on value, but a very low lower bound on value, such that most options are close to the upper bound and far away from the lower bound, would count as a Top-Heavy moral theory.

The fourth type of theory is *Bottom-Heavy*. These are simply the reverse of Top-Heavy theories.

We can represent these theories visually, where horizontal lines represent different options, which are connected by a vertical line, representing the choice-worthiness function. The greater the distance between the two horizontal lines, the greater the difference in choice-worthiness between those

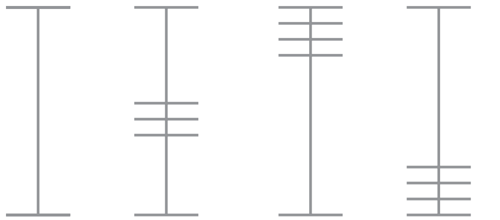

Figure 4.1

two options. If we used PEMT, the four theories would look as follows (see Figure 4.1).

When comparing Top-Heavy and Bottom-Heavy, the PEMT yields the intuitively right result. Top-Heavy and Bottom-Heavy are simply inversions of each other, so it seems very plausible that one should treat the size of choice-worthiness differences as the same according to both theories, just of opposite sign.

For Bipolar and outlier, however, the PEMT does not yield the intuitively right result. Because it *only* cares about the maximal and minimal values of choice-worthiness, it is insensitive to how choice-worthiness is distributed among options that are not maximally or minimally choiceworthy. This means that Bipolar theories have much more power, relative to outlier theories, than they should.

This might not be immediately obvious, so let us consider a concrete case. Suppose that Sophie is uncertain between an absolutist moral theory (Bipolar), and a form of utilitarianism that has an upper limit of value of saving 10 billion lives, and a lower limit of forcing 10 billion people to live lives of agony (outlier), and suppose that those views are incomparable with each other. She has 1% credence in the absolutist theory, and 99% credence in bounded utilitarianism. If the PEMT normalization is correct, then in almost every decision-situation she faces she ought to side with the absolutist theory. Let's suppose she is confronted with a murderer at her door, and she could lie in order to save her family: an action required by utilitarianism, but absolutely wrong according to the absolutist view. Given the PEMT, it's as bad to lie, according to the absolutist view, as it is to force 10 billion people to live lives of agony, according to utilitarianism. So her 1% credence in the absolutist view means that she shouldn't lie to the murderer at the door. In fact, she shouldn't lie even if her credence was as low as 0.000001%. That seems incredible. The PEMT is supposed to be motivated by the idea of giving each moral theory 'equal say', but it fails to do this in cases where some theories put almost all options into just two categories.

For a second illustration of how other accounts can fail to respect the principle of 'equal say', giving undue influence to some theories over others, consider the max-mean principle. Taking our four theories described above, it would normalize them such that they would be represented as follows (see Figure 4.2), where to 'normalize' two theories is to give them a shared fixed unit of choice-worthiness.

That is, max-mean favours Top-Heavy theories and punishes bottom-heavy theories. It's clear, therefore, that max-mean does not deal even-handedly between these two classes of theories. Exactly analogous arguments apply to mean-min.

What, though, of variance voting If we treat the variance of choice-worthiness as the same across all four theories, they would be represented as follows (see Figure 4.3).

Because Top-Heavy and Bottom-Heavy are inverses of one another, they have the same variance. So, on variance voting, the magnitudes of choice-worthiness differences between options are treated as the same, only opposite in sign. This is the result we wanted, doing better than max-mean or mean-min. But it also does better than the PEMT in terms of how it treats Bipolar compared with outlier: because Bipolar places most of its options at

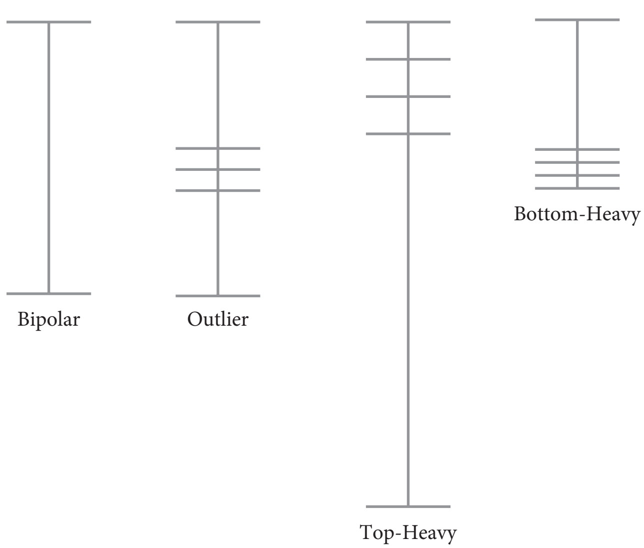

Figure 4.2

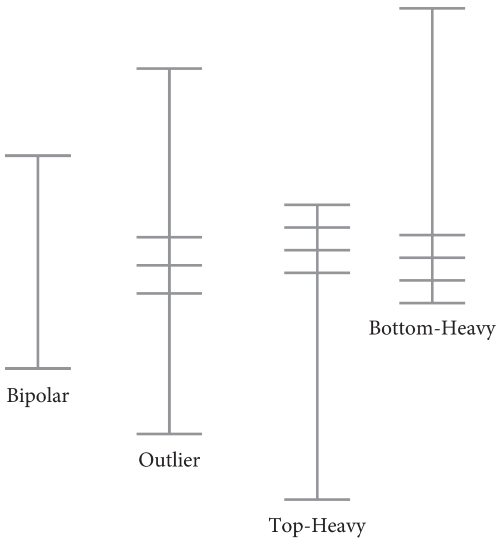

Figure 4.3

the top or bottom of its choice-worthiness function, in order to make the variance equal with outlier, its range must be comparatively smaller than outlier. Again, that was the result we wanted. So the consideration of particular cases seems to motivate variance over its rivals.

These examples are suggestive, but hardly constitute a knockdown argument. Perhaps there are other voting methods that do as well as variance does on the cases above. Perhaps there are other cases in which variance does worse than the other methods we've mentioned. So it would be nice to provide a more rigorous argument in favour of variance. The next two sections do exactly that. We'll suggest two different ways of making the idea of 'equal say' formally precise. We find the second precisification more compelling, but we show that, either way, normalizing at 'equal say' means normalizing at variance. In so doing, we thereby produce a non-arbitrary justification for normalizing at variance rather than the range or any other features of a theory's choice-worthiness functions: variance voting is the normalization that best captures the principle of 'equal say'.[16]

[16] The following two sections draw very heavily on two results within social choice theory that can be found in Owen Cotton-Barratt, 'Geometric Reasons for Normalising Variance to Aggregate Preferences', unpublished MS, http://users.ox.ac.uk/%7Eball1714/Variance%20normalisation.pdf. These results were initially motivated by the problem of moral uncertainty, arising out of conversation between us, though we had very little input on the proofs. However, they are interesting results within social choice theory, too. We state the arguments informally here; for the full proofs, see the paper.

IV. Two Arguments for Variance Voting

Distance from the Uniform Theory

Consider a uniform choice-worthiness function—one that assigns the same degree of choice-worthiness to all options. If any theory's choice-worthiness function were normalized to be essentially uniform before applying MEC,[17] then that theory would not affect the final decision. Such a normalization would give that theory no 'say'. We could thus measure how much 'say' a theory has by how 'far away' its normalized choice-worthiness function is from the uniform choice-worthiness function. Remember that by 'say' we are thinking of the degree to which the theory may influence the choice between options, for a fixed degree of credence in that theory.

Imagine starting each theory off with a uniform choice-worthiness function and an equal amount of credit, where this credit can be spent on moving the choice-worthiness function away from the uniform function. Every move away from the uniform choice-worthiness assignment increases the 'say' of that theory, and uses up a proportionate amount of credit. On this account, giving every theory 'equal say' means giving them an equal amount of starting credit. In this section, we will spell out this suggestion, explain the motivation for it, and demonstrate that variance voting is the only normalization method that gives every theory 'equal say', so understood.

Let us begin by considering different theories that *are* intertheoretically comparable. It should be clear that a completely uniform theory, according to which all options are equally choiceworthy, has no 'say' at all: it never affects what it's appropriate to do. We'll say that it gives all options choice-worthiness 0, though we could have just as well said it gives all options 17, or any other number. Next, consider a theory, T_1, which differs from the uniform theory only insofar as its choice-worthiness function gives one option, A, a different choice-worthiness, x. There are two ways in which a theory T_2 might have more 'say' than T_1. First, it could have the same choice-worthiness ordering as T_1, but its choice-worthiness function could give A a higher numerical value (remembering that, because we are talking about theories that are intertheoretically comparable, this is a meaningful

[17] If a theory is represented by a choice-worthiness function f, it is also represented by $0.1f$, $0.01f$, $0.001f$, and so on. These limit to a uniform choice-worthiness function, and if we are far enough down the sequence then the representative will be close enough to uniform to make no difference.

difference between these two theories). If it gave A a numerical value of $2x$, so that the choice-worthiness difference between A and any other option is twice as great according to T_2 than according to T_1, then T_2 would have twice as much 'say' as T_1. A second way in which a theory could have more 'say' than T_1 is if it assigned non-zero numerical values to another option in addition to A. Then it would have 'equal say' with respect to A, but would have a greater 'say' with respect to the other options.

But what does 'moving away' from the uniform theory mean? We can take this idea beyond metaphor by thinking of choice-worthiness functions geometrically. To see this, suppose that there are only two possible options, A and B, and three theories, T_1, T_2 and T_3, whose choice-worthiness functions are represented by Table 4.3.

Using the choice-worthiness of A as the x-axis and the choice-worthiness of B as the y-axis, we may represent this geometrically as follows (see Figure 4.4).

Any point on this graph represents some choice-worthiness function and those corresponding to T_1, T_2 and T_3 are marked. The diagonal line represents all the uniform choice-worthiness functions. The dotted lines show the distance from each of T_1, T_2 and T_3 to their nearest uniform choice-worthiness function. These distances allow a way of precisely defining 'equal say'. Giving each theory 'equal say' means choosing a (normalized) choice-worthiness function for each theory such that, for every choice-worthiness function, the distance from that choice-worthiness function to the nearest uniform choice-worthiness function is the same.

It turns out that the distance from a choice-worthiness function to the nearest uniform function is always equal to the standard deviation of the distribution of choice-worthiness values it assigns to the available options.[18] So treating all choice-worthiness functions as having 'equal say' means treating them as lying at the same distance from the uniform function, which means treating them such that they have the same standard deviation and thus the same variance. variance voting is thus the unique

Table 4.3

	T_1	T_2	T_3
A	−4	3	4
B	1	4	1

[18] Proof of this is given in Cotton-Barratt, 'Geometric Reasons for Normalising Variance to Aggregate Preferences'.

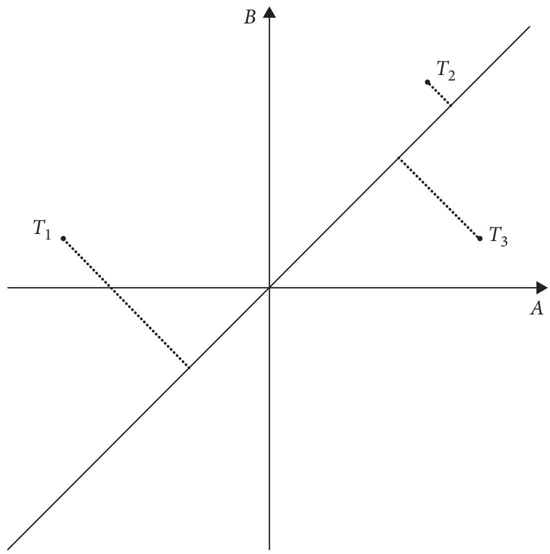

Figure 4.4

normalization method for preserving 'equal say' on this understanding of 'equal say'.

We can now look at the geometric interpretation of normalizing theories by their variance (see Figure 4.5).

The dashed lines in this diagram represent all the choice-worthiness functions that are distance of 1 from the nearest uniform function.[19] This means that they also have a standard deviation of 1 and hence a variance of 1. In order to normalize each theory so that they have the same amount of 'say', we move each theory to the closest point on one of the dashed lines (the arrows show these moves). This corresponds to linearly rescaling all of the theory's choice-worthiness values so that their variance is equal to 1, while keeping their means unchanged. This doesn't change the ordering of the options by that theory's lights, it just compresses it or stretches it so that it has the same variance as the others. One can then apply MEC to these normalized choice-worthiness functions.

This all works in the same way for any finite number of options.[20] A choice-worthiness function gives an assignment of a real number to each

[19] We could have chosen any non-zero value here, but 1 is especially convenient.

[20] This argument applies only in the case where there are finitely many options, and makes an assumption of symmetry in the weight we attach to each. This is the simplest case for inter-theoretic value comparisons, and any method should at least behave well in this base case.

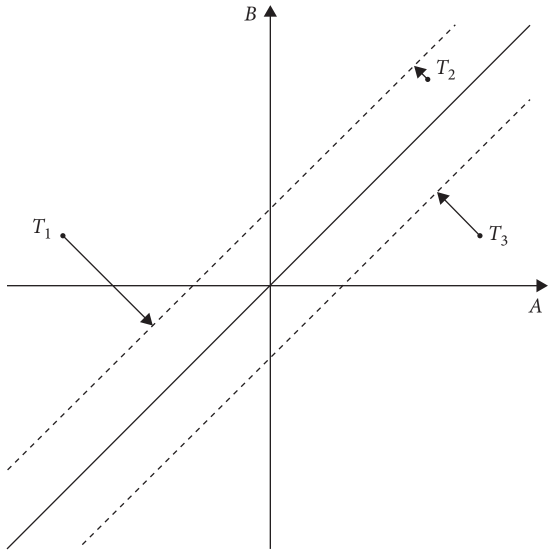

Figure 4.5

option, so if there are *n* options, a choice-worthiness function can be represented as a collection of *n* real numbers. Just as pairs of real numbers give us Cartesian coordinates in the plane and triples give us coordinates in three-dimensional space, so we can interpret this collection as the coordinates of a point in *n*-dimensional Euclidean space. We can then proceed the same way, looking at the distance in this *n*-dimensional space from a choice-worthiness function to the nearest uniform theory, equating this to 'say', and normalizing to make the distances the same. Just as before, the distance corresponds to the standard deviation, and so normalizing to equalize variance is the unique way to provide 'equal say'.

While there is no need to normalize the means of the choice-worthiness functions (it does not affect the MEC calculation, as we are ultimately interested in comparing between options) it could be convenient to normalize them all to zero, by adding or subtracting a constant from each choice-worthiness function. If so, then the choice-worthiness functions are in the familiar form of 'standard scores' or 'z-scores' where the mean is zero and the unit is one standard-deviation. These z-scores are commonly used in statistics as a way to compare quantities that are not directly comparable, so it is particularly interesting that our approach to intertheoretic choice-worthiness comparisons for non-comparable theories could be summarized as 'compare them via their z-scores'.

The Expected Choice-Worthiness of Voting

The previous argument cashed out the idea of 'equal say' as 'equal distance from a uniform choice-worthiness function'. For our second argument, we shall borrow a concept from voting theory: *voting power*. An individual's voting power is the a priori likelihood of her vote being decisive in an election, given the assumption that all the possible ways for other people to vote are equally likely. It is normally used for elections with just two candidates, but the concept is perfectly general.

We shall extend this concept to flesh out 'equal say'. A first challenge is that while voters all have just one vote, theories come with different credences. We want theories with the same credence to have the same voting power and for voting power to go up on average as the credence increases.[21] We can resolve this by looking at the voting power of a small increase in the credence of a particular theory.

A second challenge is that by a theory's own lights it doesn't just matter that one's credence in it is decisive in determining which option gets chosen, it matters how much better this chosen option is than the option that would have been chosen otherwise. Getting its way in a decision about whether to prick someone with a pin matters a lot less, for utilitarianism, than getting its way in a decision about whether to let a million people die. If we are normalizing to provide 'equal say', we should take that into account as well. Since theories come with a measure of this difference between the options (the choice-worthiness difference), and they use its expectation when considering descriptive uncertainty, it is natural to use this here. This means we should speak not just of the likelihood of being decisive, but of the increase in expected choice-worthiness. We thus achieve 'equal say' when, from a position of complete uncertainty about how our credence will be divided over different choice-worthiness functions, an increase in our credence in a theory by a tiny amount will increase the expected choice-worthiness of the decision by the same degree regardless of which theory it was whose credence was increased.

There is one final challenge. If each theory had one canonical choice-worthiness function, this definition would work. But since each theory is described by infinitely many different choice-worthiness functions (positive

[21] The qualification 'on average' is needed as it is possible for a theory to get its way all the time when it is given a credence that is slightly less than 1 and from that point increases in credence will not improve its power. This is analogous to how a voting block might already have all the power with less than 100% of the votes.

affine transformations of each other), we do not yet know which choice-worthiness function to use to represent each theory and so cannot come up with a unique value for the 'expected choice-worthiness'.

However, we can resolve this by considering that the normalization used to choose an option in a decision situation should be the same normalization used to measure 'equal say' in terms of this version of voting power. This doesn't sound like a strong constraint, but it is enough to let us prove that there is a unique normalization method that satisfies it and equalizes voting power.[22]

Given that we have found two independently plausible ways of cashing out the principle of 'equal say' that both lead to the same conclusion, we think it is warranted to think of variance voting as strongly supported by that principle. We'll now turn to discuss two issues regarding how to precisely formulate variance voting.

V. Option-Individuation and Measure

An objection that one can make to both the Borda Rule and to variance voting is that they are both extremely sensitive to how one individuates options.[23] To illustrate this with respect to the Borda Rule, consider the following case.

Trolley Problems

Sophie is watching as an out-of-control train hurtles towards five people working on the train track. If she flips a switch, she will redirect the train, killing one person working on a different track. Alternatively, she could push a large man onto the track, killing him but stopping the train. Or she could do nothing. So she has three options available to her.

A: Do nothing.
B: Flick the switch.
C: Push the large man.

She has credence in three moral theories.

[22] See Cotton-Barratt, 'Geometric Reasons for Normalising Variance to Aggregate Preferences'.

[23] This problem is analogous to the problem of 'clone-dependence' in voting theory, which itself is a generalization of the idea of vote-splitting. For discussion of clone-dependence, see Tideman, 'Independence of Clones as a Criterion for Voting Rules'. We thank Graham Oddie for pressing this criticism of the Borda Rule. The example is Oddie's.

40% in utilitarianism, according to which: *B*>*C*>*A*
30% in simple Kantianism, according to which: *A*>*B*~*C*
30% in sophisticated Kantianism, according to which: *A*>*B*>*C*

In this case, according to the Borda Rule, *B* is the most appropriate option, followed by *A* and then *C*.[24] But now let us suppose that there are actually two Switching options:

A: Do nothing.
B': Flick the switch to the left.
B'': Flick the switch to the right.
C: Push the large man over the railing to stop the track

Sophie has the same credences in moral theories as before. Their recommendations are as follows:

Utilitarianism: *B'*~*B''*>*C*>*A*
Simple Kantianism: *A*>*B'*~*B''*~*C*
Sophisticated Kantianism: *A*>*B'*~*B''*>*C*

Given these choice-worthiness rankings, according to the Borda Rule, *A* is the most appropriate option, then *B'* and *B''* equally, then *C*.[25] So, according to the Borda Rule, it makes a crucial difference to Sophie whether she has just one way of flicking the switch or whether she has two: and if she has two ways of flicking the switch, it's of crucial importance to her to know whether that only counts as one option or not. But that seems bizarre.

To see how this problem plays out for variance voting, suppose that there are only four possible options, all of which are available to the decision-maker, and suppose that the decision-maker has credence in only two theories (see Table 4.4).

[24] Now that some theories posit tied options, we return to using our 'official' definition of a Borda Score in our working. Option *A* receives a score of $0+30\times2+30\times2-(40\times2+0+0)=40$. B's score is $40\times2+0+30\times1-(0+30\times1+30\times1)=50$. C's score is $40\times1+0+0-(40\times1+30\times1+30\times2)=-90$.

[25] Option *A* receives a score of $0+30\times3+30\times3-(40\times3+0+0)=60$. *B'* and *B''* each receive a score of $40\times2+0+30\times1-(0+30\times1+30\times1)=50$. C's score is $40\times1+0+0-(40\times2+30\times1+30\times3)=-160$.

Table 4.4

	T_1	T_2
A	18	10
B	16	4
C	6	4
D	0	22

Table 4.5

	T_1	T_2
A	18	10
B	16	4
C	6	4
D'	0	22
D"	0	22

These two theories have been normalized in accordance with their variance. For both T_1 and T_2, the mean choice-worthiness is 10 and the variance is 54. But now suppose that the decision-maker comes to believe that option *D* can be broken down into two distinct options, *D'* and *D"*. There is no morally relevant difference between the two options, so the decision situation now looks as in Table 4.5.

Now, the mean of T_1 is 8 and the variance is 59.2, while the mean of T_2 is 12.4 and the variance is 66.2. So the variance in T_2 is now larger than in T_1 and they would need to be renormalized. This would require compressing the distribution of choice-worthiness numbers in T_2, giving it less 'say' relative to T_1 than it had before we divided *D*. This means that the variance of a theory depends crucially on how we individuate options, which seems problematic.[26] (Note that this was not a problem for PEMT because it normalized by the range of choice-worthiness and, unlike the variance, the range of a distribution is not sensitive to how many times a number occurs in it.)

However, there is a principled and satisfying response to this objection: that we need to have a *measure* over the space of possible options, and that

[26] There is a close analogy here to the 'independence of clones' property in voting theory, whereby the outcome of an election should not be sensitive to whether a new candidate that is very similar to an existing one joins the race.

we were neglectful when we didn't initially include a measure in our definition of the Borda Rule or of variance voting[27] A measure will define the 'sizes' of different options, allowing an option to be divided into two smaller options without affecting the variance. Technically, we will use a 'probability measure': a function that assigns non-negative numbers to subsets of a set (in this case the set of all possible options), assigns 0 to the empty set, 1 to the whole set, and where the number assigned to the union of two disjoint sets is the sum of the numbers assigned to each of the smaller sets. Note that this does *not* necessarily mean we're talking about the decision-maker's credences in the likelihood of different options; the term 'probability measure' simply signifies that the whole set of options is assigned measure 1.

A way to visualize the idea of a measure is to think of the entirety of the space of possibilities as the area of a two-dimensional shape. When we talk about an 'option' we are talking about some area within the shape. What a measure does is give sense to the intuitive idea of the size of the space of possibilities, and so gives us the resources to say that one option takes up twice as much of the space as another, or a specified fraction of the whole space.

With the concept of a measure on board, we can reformulate the definition of an option's Borda Score as follows: that an option's *Borda Score* is equal to the sum on the measure of the options below it minus the sum of the measure of the options above it. Once we've defined a Borda Score in this way, then we can use all the other definitions as stated. Nothing will change in terms of its recommendations in the cases we've previously discussed. But it resolves the option-individuation problem.

To see how this resolves the option-individuation problem, consider again the case given above. Let us suppose that the measure of each option, *A*, *B* and *C*, is 1/3.[28] If so, then, as before, according to the Borda Rule, *B* is the most appropriate option, followed by *A* and then *C*.[29] Now, however, when we split the option *B* into options *B'* and *B"*, we have to also split the measure: let us suppose that the measure splits equally, so that *B'* and *B"* each have measure 1/6.[30] If so, then according to the Borda Rule, *B* is still

[27] We thank Owen Cotton-Barratt for this suggestion.

[28] Note that there would be no difference to our argument if the measure were split unequally among options *A, B,* and *C*.

[29] Option *A* receives a score of $0+30\times(2/3)+30\times(2/3)-(40\times(2/3)+0+0)=13\ 1/3$. B's score is $40\times(2/3)+0+30\times(1/3)-(0+30\times(1/3)+30\times(1/3))=16\ 2/3$. C's score is $40\times(1/3)+0+0$ $(40\times(1/3)+30\times(1/3)+30\times(2/3))=-30$.

[30] Note that there would be no difference to our argument if the measure did not divide evenly between *B'* and *B"*.

the most appropriate option, followed by *A* and then *C*.[31] In general, the addition of a measure means that we can make sense of a 'size' of an option, and will therefore avoid the option-individuation problem.

Similarly, once we use a measure, the problem of variance voting's dependence on how we individuate options dissolves: we normalize the variance of the distribution of choice-worthiness by taking the choice-worthiness of each option weighted by that option's measure. So, let us suppose that each of the options *A*–*D* had measure 1/4. In this case, as before, in the first decision-situation the mean of T_1 is 10 and the variance is 54. However, this stays the same in the second decision-situation. When we split *D* into the smaller options *D*' and *D*", the measure is split, too. Let's suppose, then, that each new option gets measure 1/8 (though the argument would work just as well if the measure was split unequally). If so, then the mean and variance of both T_1 and T_2 is the same in the second decision-situation as it is in the first decision-situation. And that's exactly the result we wanted.[32]

There are additional benefits to the incorporation of a measure. First, it means that the Borda Rule can handle situations in which the decision-maker faces an infinite number of options.[33] Before we had defined a measure over possibility space and incorporated that into an option's Borda Score, one could have objected that the Borda Rule can't handle infinite option sets. For, if the number of options below or above one option *A* were infinite, then there would be no answer to the question of what that option's Borda Score is.

Having a measure over possibility space resolves this problem, because one can have an infinite number of options with a measure that sums to some finite number. For example, suppose that below option *x* there are an infinite number of options, with measure 1/4, 1/8, 1/16, 1/32... In this case, even though there are an infinite number of options there is a fact about the sum of the measure of options below *A*: namely, 1/2. Indeed, because

[31] Option *A* receives a score of $0 + 30 \times (2/3) + 30 \times (2/3) - (40 \times (2/3) + 0 + 0) = 13\ 1/3$. *B*' and *B*" each receive a score of $40 \times (2/3) + 0 + 30 \times (1/3) - (0 + 30 \times (1/3) + 30 \times (1/3)) = 16\ 2/3$. C's score is $40 \times (1/3) + 0 + 0 - (40 \times (1/3) + 30 \times (1/3) + 30 \times (2/3)) = -30$. That is, the scores are just the same as they were prior to the more fine-grained individuation of option *B*.

[32] In Chapter 2, we criticized My Favorite Theory in part because of the problem of theory-individuation. One might wonder: if both our account and MFT have individuation problems, doesn't this undermine our earlier objection? However, the use of measure gives us a principled solution to this problem, whereas we cannot see a way of using a measure to solve the theory-individuation problem. So we think our earlier objection still stands.

[33] We thank an anonymous reviewer at *Mind* for pressing this objection.

the measure of the set of all possible options is 1, the measure of options above or below any particular action will always be finite. So the Borda Score of an option will always be well-defined, even when there are an infinite number of options available to the decision-maker.

Second, it means that variance voting will avoid a problem that faces other structural accounts. Many moral theories are often unbounded: theories according to which there can be situations where there is no maximally choiceworthy or no minimally choiceworthy option. For example, any theory that accepts the Total View of population ethics is unbounded above and below: one can keep making a world better by adding to it additional happy people; and one can keep making a world worse by adding to it lives that aren't worth living. Sepielli objects that the PEMT has nothing to say concerning how to normalize such unbounded moral theories in situations where there is no best or worst option. A very similar problem afflicts max-mean and mean-min.

We take it as a virtue of variance voting that it is able, once we have incorporated the idea of a measure, to normalize many unbounded theories. Just as unbounded distributions can have a mean (if the chance of getting an extreme value falls off quickly enough compared to the growth of the extreme values), so too can they have a variance.

So we now have the resources to state variance voting precisely. Because of the arguments we have given, we propose that, in conditions of moral uncertainty and intertheoretic incomparability, decision-makers should choose the option with the highest expected choice-worthiness, where the (measure-weighted) variance of choice-worthiness should be treated as the same across all considered theories.

VI. Broad vs Narrow

For both the Borda Rule and variance voting we have a choice about how to define the theory. When we normalize different theories at their variance, should we look at the variance of choice-worthiness over all possible options, or the variance of choice-worthiness merely over all the options available to the decision-maker in a given decision situation? Similarly, when we say that an option's Borda Score, on a given theory, is the sum of the measure of the options ranked lower than it by the theory minus the sum of the measure of the options ranked higher than it, should we sum over all the options in a given decision-situation, or should we sum over all conceivable options? These two approaches will give very different answers concerning what it's appropriate to do in a given situation. Following Sen, we will say that Broad accounts are defined across all conceivable options and that Narrow accounts are defined over only the options in a particular decision-situation.[34] In this section we argue that Narrow is the best approach, though we are not confident.

We have two key reasons for preferring Narrow accounts. First, Narrow accounts are able to provide a principled solution to the infectious incomparability problem, in a way that Broad accounts are not—we discuss this further in Chapter 5. Second, Narrow accounts are more action-guiding. For example, if you use the Broad Borda Rule, then, for any option you face, you'll have simply no idea what Borda Score it should receive—we would need to know the total measure of all options above and below the option in question, and that seems very difficult or impossible. We could do it approximately if we could know in what *percentile* the option ranks among all possible options—but how are we meant to know even that? Similar difficulties plague variance voting. In contrast, you can come to at least a rough approximation of the options facing you in a particular decision-situation. So we are able to actually use Narrow methods, at least approximately.

However, there are arguments against Narrow accounts. In his extensive criticism of Lockhart's PEMT, Sepielli gives four arguments against the PEMT that arise in virtue of the fact that it makes intertheoretic comparisons only within a decision-situation (rather than across all decision situations).[35] One might therefore think that our account will also be susceptible to these arguments.

His first two arguments are as follows. First, he argues that the Narrow PEMT cannot make sense of the idea that some decision-situations are higher-stakes for some theories than for others. Second, he argues that the PEMT generates inconsistent choice-worthiness comparisons: in one decision-situation, the difference in choice-worthiness between A and B, on T_1 is the same as the difference in choice-worthiness between A and B on T_2, but in another decision-situation the difference in choice-worthiness between A and B, on T_1 is larger than the difference in choice-worthiness between A and B on T_2

[34] The terminology of 'broad' and 'narrow' for this distinction comes from Amartya Sen, *Choice, Welfare, and Measurement*, Cambridge, MA: MIT Press, 1982, p. 186.

[35] Sepielli, 'Normative Uncertainty for Non-Cognitivists'.

However, in the context of our project, these criticisms lose their force. First, we are using the Borda Rule and variance voting not as accounts of how theories actually compare, but as a way of coming to a principled decision in the face of incomparable theories. So there isn't a fact of the matter about some decision-situations being higher stakes for some of these theories rather than others. And these accounts aren't generating inconsistent assignments of choice-worthiness, because they aren't pretending to make claims about how choice-worthiness actually compares across theories. Rather, they are simply giving an account of what it's appropriate to do given that choice-worthiness doesn't compare across theories.

A separate argument against Narrow Borda and Narrow Variance accounts is that they violate *Contraction Consistency*.

Contraction Consistency: Let $\mathcal{M}$ be the set of maximally appropriate options given an option-set $\mathcal{A}$, and let $\mathcal{A}'$ be a subset of $\mathcal{A}$ that contains all the members of $\mathcal{M}$. The set $\mathcal{M}'$ of the maximally appropriate options given the reduced option-set $\mathcal{A}'$ has all and only the same members as $\mathcal{M}$.

For simplicity, we'll just focus on this criticism as aimed at the Narrow Borda Rule, but just the same considerations would apply to Narrow variance voting.

To see that the Borda Rule violates *Contraction Consistency*, consider again the *Hiring Decision* case.

Hiring Decision

Jason is a manager at a large sales company. He has to make a new hire, and he has three candidates to choose from. They each have very different attributes, and he's not sure what attributes are morally relevant to his decision. In terms of qualifications for the role, applicant *B* is best, then applicant *C*, then applicant *A*. However, he's not certain whether that's the only relevant consideration. Applicant *A* is a single mother, with no other options for work. Applicant *B* is a recent university graduate with a strong CV from a privileged background. And applicant *C* is a young black male from a poor background, but with other work options. Jason has credence in three competing views.

30% credence in a form of virtue theory. On this view, hiring the single mother would be the compassionate thing to do, and hiring simply on the basis of positive discrimination would be disrespectful. So, according to this view, $A>B>C$.

30% credence in a form of non-consequentialism. On this view, Jason should just choose in accordance with qualification for the role. According to this view, *B*>*C*>*A*.

40% credence in a form of consequentialism. On this view, Jason should just choose so as to maximize societal benefit. According to this view, *C*>*A*>*B*.

As we noted, *C* is, both intuitively and according to the Borda Rule, the uniquely most appropriate option. Now, however, suppose that it were no longer possible to hire candidate *A*. In which case, Jason's credence distribution would look as follows.

30% credence in virtue theory, according to which *B*>*C*.
30% credence in non-consequentialism, according to which *B*>*C*.
40% credence in consequentialism, according to which *C*>*B*.

In this new decision-situation, *B* is now the uniquely most appropriate option. The appropriateness of options is highly sensitive to which other options are within the option-set.

How strong of an objection to Narrow accounts is the violation of *Contraction Consistency*? We're not sure. We think it would be reasonable if one found this violation to be compelling, and therefore wanted to endorse a Broad account, despite Broad accounts' other problems. But, on balance, we think that those other problems are more grave, because we think that the two primary reasons one might have for endorsing *Contraction Consistency* are not compelling in this case.

First, one might worry that violation of *Contraction Consistency* would lead one to be open to money-pumps, choosing *B* over *A*, *C* over *B*, and *A*' (a strictly worse option than *A*) over *C*. But such arguments are of dubious cogency. Though we don't have space in this book to delve into the extensive literature around money-pumps, we point the reader to some compelling recent work arguing that agents with cyclical preferences across choice-situations are not vulnerable to money-pumps.[36]

Second, a reason why *Contraction Consistency* is thought desirable in the voting context is that violating it leads to susceptibility to tactical voting. Again, consider *Hiring Decision*. If the virtue theory could pretend that its

[36] See Arif Ahmed, 'Exploiting Cyclic Preference', *Mind*, vol. 126, no. 504 (October 2017), pp. 975–1022.

preference ordering was $B>A>C$ rather than $A>B>C$, then it could guarantee that its second-favoured option would 'win', rather than its least-favoured option. And, indeed, the Borda Rule is often dismissed for being extremely susceptible to tactical voting. However, as we have noted, while tactical voting is a real problem when it comes to aggregating the stated preferences of people, it is no problem at all in the context of decision-making under moral uncertainty. Theories aren't agents, and so there's no way that they can conceal their choice-worthiness ordering. If a decision-maker pretends that one theory's choice-worthiness ordering is different than it, in fact, is, she deceives only herself.

So we think there are some positive reasons in favour of our account being Narrow, and that the arguments against Narrow accounts are not strong. So we tentatively conclude that the Narrow version of our account is to be preferred.

VII. How to Act in Varying Informational Conditions

In Chapter 2, we discussed how to take moral uncertainty into account in conditions where theories' choice-worthiness is interval-scale measurable and intertheoretically comparable. In Chapter 3, we discussed how to take moral uncertainty into account in conditions where theories give merely ordinal choice-worthiness. And, earlier in this chapter, we discussed how to take moral uncertainty into account in conditions where theories give interval-scale measurable choice-worthiness but are intertheoretically incomparable. But how should we put these different criteria together? In accordance with our *information-sensitive* view, we want our account to take into account all the relevant information that theories provide to us, but not to demand more of theories than they can provide.

One natural approach takes the form of multi-step procedure: doing what you can with the most informationally rich theories, then falling back to more general techniques to fold in theories which provide less and less information.[37] The idea is as follows. At the first step, aggregate each set of interval-scale measurable and mutually intertheoretically comparable theories. For each set, you produce a new choice-worthiness function R_i, where

[37] The following is very similar to the account one of us defended in William MacAskill, 'How to Act Appropriately in the Face of Moral Uncertainty', BPhil thesis, University of Oxford, 2010.

R_i assigns numbers to options that represent each option's expected choice-worthiness (given the theories in that set). R_i is given a weight equal to the sum total of the credence of all the theories within the set. At the second step, you use variance voting to aggregate all the new choice-worthiness functions (the R_i) with every interval-scale measurable but non-comparable choice-worthiness function, producing another new choice-worthiness function *S*. *S* is weighted by the sum of the decision-maker's credences in all interval-scale theories. Then, at the third and final stage, you aggregate *S* and all merely ordinal theories using the Borda Rule.

However, that proposal suffers from the following significant problem. Consider a decision-maker with the following credence distribution:[38]

4/9 credence in T_1: $A>B>C$.
2/9 credence in T_2: $CW_2(A)=20$, $CW_2(B)=10$, $CW_2(C)=0$.
3/9 credence in T_3: $CW_3(A)=0$, $CW_3(B)=10$, $CW_3(C)=20$.

T_1 is merely ordinal, while T_2 and T_3 are interval-scale and comparable. If we use the multi-step procedure, then at the first step, we aggregate T_2 and T_3 to get the following output ordering.

5/9 credence in R_1: $C>B>A$

At the second step, we aggregate T_1 and R_1 using the Borda Rule, which gives option *C* as the winner. However, this seems like the wrong result. In particular, consider the following credence distribution.

4/7 credence in T_1: $A>B>C$.
0 credence in T_2: $CW_2(A)=20$, $CW_2(B)=10$, $CW_2(C)=0$.
3/7 credence in T_3: $CW_3(A)=0$, $CW_3(B)=10$, $CW_3(C)=20$.

In this decision-situation, using the multi-step procedure would give *A* as the most appropriate option. So having *lower* credence in T_2 makes the appropriateness ordering better by the lights of T_2. This means that the multi-step procedure violates the *Updating* condition given in Chapter 3. The reason this happens is because, in the first decision-situation, though T_2's and T_3's choice-worthiness orderings cancel out to some extent, the multi-step procedure washes this fact out when it pits the aggregated

[38] The possibility of such a problem was first suggested to us by Owen Cotton-Barratt.

ordering R_1 against the ordinal theory T_1. We consider this violation of *Updating Consistency* to be a serious problem for the multi-step procedure.

In private communication, Christian Tarsney has argued that even a single-step procedure will violate *Updating Consistency*. Consider the following variant on the previous case. Suppose that the decision-maker has three options available to her, positive credence in two (cardinal and incomparable) normative theories, T_1 and T_2, and positive credence in two descriptive states of the world, S_1 and S_2. T_1 assigns the same degrees of choice-worthiness to each option regardless of the state of the world but, according to T_2, the choice-worthiness of A and C depends on the state of the world. Here's the credence distribution.

4/9 credence in T_1: $CW_1(A) = 20$, $CW_1(B) = 10$, $CW_1(C) = 0$.
2/9 credence in T_2 & S_1: $CW_{2/1}(A) = 20$, $CW_{2/1}(B) = 10$, $CW_{2/1}(C) = 0$.
3/9 credence in T_2 & S_2: $CW_{2/2}(A) = 0$, $CW_{2/2}(B) = 10$, $CW_{2/2}(C) = 20$.

On this credence distribution, if we normalize at each moral theory's ranking of options in terms of their expected choice-worthiness, then we will also violate *Updating Consistency*. The expected choice-worthiness of options on T_2 is $CW_2(A) = 8$, $CW_2(B) = 10$, $CW_2(C) = 12$, so when we normalize T_1 and T_2 at their variance, C comes out as the most appropriate option. However, if the decision-maker then reduces her credence in S_1 to 0, distributing her credence proportionally among T_1 & S_2 and T_2 & S_2 (such that she has 4/7 credence in T_1 and 3/7 credence in T_2 & S_2), then, using the same procedure as before, A will come out as the most appropriate option. But that means that A has become more appropriate in virtue of becoming less confident in a view (namely, T_2 & S_1) on which A is the top option. We've therefore violated updating consistency.

Now, insofar as we have assumed descriptive certainty in this book, strictly speaking this problem does not arise for us. However, it is clearly a problem that needs to be addressed.

We are not confident about what is the best way to do so, but our currently favoured response is to construe our account as taking an expectation over both empirical and normative states of the world jointly, rather than over empirical-belief-relative orderings. That is: we make a hard distinction between moral theories, which order outcomes in terms of choice-worthiness, and a theory of rationality, which tells us what to do in conditions of either empirical or normative uncertainty or both. This has the disadvantage that we cannot accommodate uncertainty over normative theories that do not endorse expected utility theory. Even if a moral theory endorsed maximin or some other procedure for decision-making in the face of uncertainty, we would still aggregate empirical uncertainty, conditional on that theory, in an expectational way. One could argue that this is therefore not being appropriately responsive to the decision-maker's true uncertainty across different moral views.

In response, we note that, as argued in Chapter 1, we have to go externalist somewhere. We consider norms that govern decision-making in the face of uncertainty to be norms of rationality, and are inclined to endorse strict liability when it comes to norms of rationality. So the way we should really understand a non-expectational moral theory is as the conjunction of a moral theory (which assigns choice-worthiness to outcomes) and a theory of rationality (which, in this case, happens to be non-expectational). Our account is not sensitive to uncertainty about rationality, in which case the fact that our account 'overrides' the decision-maker's credence in a view that endorses some non-expectational decision theory should not be surprising to us.

We, therefore, very tentatively endorse a one-step theory. What we suggest is that we should normalize the Borda Scores of the ordinal theories with choice-worthiness functions by treating the variance of the interval-scale theories' choice-worthiness functions and the variance of the ordinal theories' Borda Scores as the same.[39] Of course, we are not claiming that these normalized Borda Scores represent choice-worthiness on these theories; to say that would be to pretend that ordinal theories are really cardinal. All we are suggesting is that this might be the correct way of aggregating our moral uncertainty in varying informational conditions.

If we take this approach, we need to be careful when we are normalizing Borda Scores with other theories. We can't normalize all individual comparable theories with non-comparable theories at their variance. If we were to

[39] Doing this does not alter the Borda Rule as presented in Chapter 2 when each theory has a strict choice-worthiness ordering over options. However, it *does* make a difference when some theories rate some options as equally choiceworthy to one another (which is discussed in Chapter 3, footnote 26). The account given in the previous chapter gives the standard way of dealing with ties under the Borda Rule. But when taking the variance of each theory's Borda Scores to be the same, a theory that ranks $A{\sim}B{>}C{\sim}D$ will weigh comparatively more heavily against $D{>}C{>}B{>}A$ than it would under the account we stated in the previous chapter. However, the standard way of giving Borda Scores to tied options is typically defended with recourse to something like the principle of 'equal say', and implicitly invokes average-distance-to-the-mean as the correct account of 'equal say'. Now that we have seen that normalizing at the variance is the best account of 'equal say', we should use the method of dealing with tied options that normalizes at the variance, rather than at distance to the mean.

do so, we would soon find our equalization of choice-worthiness-differences to be inconsistent with each other. Rather, for every set of interval-scale theories that are comparable with each other, we should treat the variance of the choice-worthiness values of all options on that set's common scale as the same as the variance of every individual non-comparable theory.

An example helps to illustrate the proposal. Consider four theories, T_1–T_4, in order from left to right (see Figure 4.6).

T_1 is a merely ordinal theory. The diagram illustrates the Borda Scores that T_1 assigns to options. T_2 is interval-scale measurable but is not comparable with any other theory. T_3 and T_4 are interval-scale measurable and comparable with each other. What the single-step procedure does is to treat the variance of T_1's Borda Scores as equal with the variance of T_2's choice-worthiness function and as equal with the variance of the choice-worthiness of options across *both* T_3 and T_4. As should be clear from the diagram, the variance of T_3 is smaller than the variance of T_4. But if T_3 and T_4's variances were each individually normalized with T_2, then the variance of T_3 and T_4 would be the same. So we should not normalize T_3 and T_4 individually with T_2. Rather, it's the variance of the distribution of choice-worthiness on T_3 and T_4's common scale that we treat as equal with other theories.

With their variances treated as equal in the correct way, the theories would look approximately as in Figure 4.7.

Then, once we have done this, we maximize the expectation of the normalized scores given to each option by each theory.

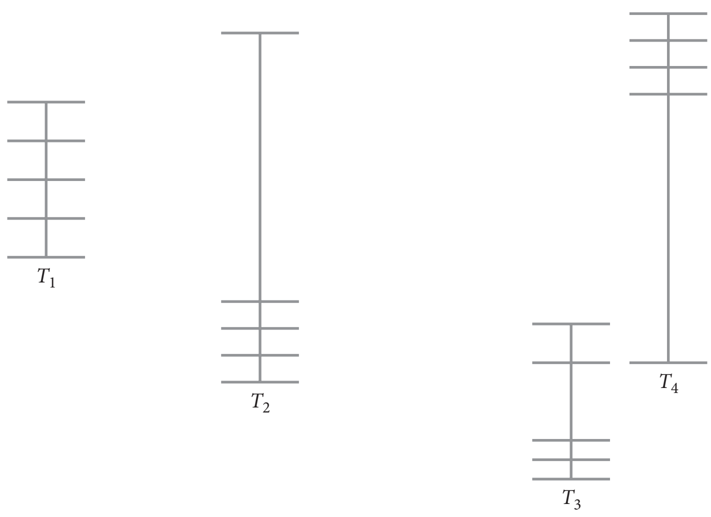

Figure 4.6

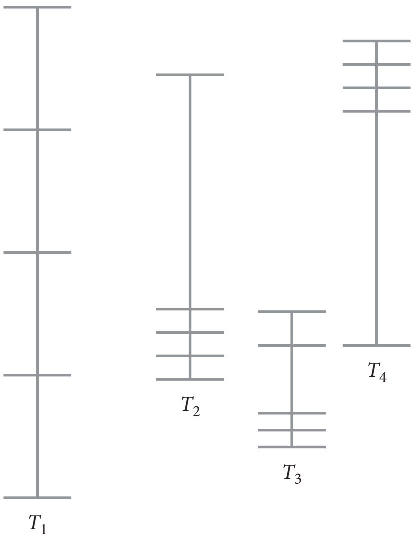

Figure 4.7

This single-step aggregation method wouldn't be possible if we didn't use a scoring function as our voting system in the situation involving merely ordinal theories. If, rather than a scoring function, we had defended a Condorcet extension as the correct way to take into account moral uncertainty over merely ordinal theories, we would be forced to endorse the multi-step procedure. But the objection to the multi-step procedure given above looks fatal. So we take this to provide additional support in favour of the use of a scoring function to aggregate merely ordinal theories, rather than a Condorcet extension.

As a final comment on this, we should note that the above account is effectively taking an expectation over all moral theories. So, even though one of the authors (William MacAskill) initially thought that the problems of merely ordinal theories and intertheoretic incomparability were reasons to reject MEC as a *general* theory, we ultimately end up with a sort of extension of MEC as a general account of decision-making under moral uncertainty, for the informational conditions we consider. Of course, we have only considered a small number of informational conditions, so it remains to be seen whether this will remain true when further work considers a wider range of informational conditions.[40]

[40] We thank Christian Tarsney for emphasizing this to us.

Conclusion

In this chapter, we considered how to take moral uncertainty into account in the situation where the decision-maker has non-zero credence in only interval-scale measurable theories that are intertheoretically incomparable. Arguing that the Borda Rule is unsatisfactory in this context, and arguing against Lockhart's PEMT among others, we argued in favour of variance voting, on the basis that it best respects the principle of 'equal say'. We then showed how one should aggregate one's uncertainty in varying informational conditions.

This concludes our account of what we believe to be the best theory for how to make decisions under moral uncertainty. However, we don't yet know much about when theories are comparable and when they are not, nor do we know what makes theories comparable, if and when they are comparable. Chapter 5 tackles these issues.