3
Ordinal Theories and the Social Choice Analogy

Introduction

In the previous chapter, we argued that when the decision-maker has non-zero credence only in theories that are interval-scale measurable and intertheoretically comparable, it's appropriate to maximize expected choiceworthiness.

But when we try to apply MEC in general, a couple of problems immediately arise. First, what should you do if one of the theories in which you have credence doesn't give sense to the idea of interval-scale measurable choice-worthiness? Some theories will tell you that murder is more seriously wrong than lying, yet will not give any way of saying that the difference in choice-worthiness between murder and lying is greater, smaller, or equally as large as the difference in choice-worthiness between lying and telling the truth. But if it doesn't make sense to talk about ratios of differences of choice-worthiness between options, according to a particular theory, then we won't be able to take an expectation over that theory. We'll call this *the problem of merely ordinal theories*.

A second problem is that, even when all theories under consideration give sense to the idea of interval-scale choice-worthiness, we need to be able to compare the size of differences in choice-worthiness between options *across* different theories. But it seems that we can't always do this. A rights-based theory claims that it would be wrong to kill one person in order to save fifty; utilitarianism claims that it would be wrong not to do so. But for which theory is there more at stake? In line with the literature, we'll call this *the problem of intertheoretic comparisons*.[1]

Some philosophers have suggested that these problems are fatal to the project of developing a normative account of decision-making under moral

[1] E.g. Lockhart, *Moral Uncertainty and Its Consequences*; Ross, 'Rejecting Ethical Deflationism'; Sepielli, 'What to Do When You Don't Know What to Do'.

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

DOI: 10.1093/oso/9780198722274.001.0001

uncertainty.[2] The primary purpose of this chapter and the next is to show that this is not the case.

We discuss these problems in more depth in section I. In section II, we introduce the analogy between decision-making under moral uncertainty and social choice, and explain how this analogy can help us to overcome these problems. The rest of the chapter is spent fleshing out how this idea can help us to develop a theory of decision-making under moral uncertainty that is applicable even when all theories under consideration are merely ordinal, and even when there is neither level- nor unit-comparability between those theories.[3] In section III, we show how the social choice analogy gives fertile ground for coming up with new accounts. We consider whether *My Favorite Theory* or *My Favorite Option* might be the right theory of decision-making under moral uncertainty in conditions of merely ordinal theories and incomparability, but reject both of these accounts. In section IV we defend the idea that, when maximizing choice-worthiness is not possible, one should use the Borda Rule instead.

Note that this chapter and the next chapter—which primarily discusses what to do in conditions of interval-scale measurability but incomparability—should ideally be considered together rather than read in isolation. The next chapter will discuss two objections to the Borda Rule—that it is sensitive to how one individuates options, and that it violates *Contraction Consistency*—and will also discuss what is the correct account of what to do in mixed informational conditions. We will suggest that the fact that the Borda Rule allows us to endorse a 'one-step' procedure for decision-making in varying informational conditions may be an additional benefit of the Borda Rule.

I. Intertheoretic Comparisons and Ordinal Theories

If you want to take an expectation over moral theories, two conditions need to hold. First, each moral theory in which you have credence needs to

[2] E.g. Gracely, 'On the Noncomparability of Judgments Made by Different Ethical Theories'; Hudson, 'Subjectivization in Ethics'; Ross, 'Rejecting Ethical Deflationism'; Gustafsson and Torpman, 'In Defence of My Favourite Theory'. In conversation with one of the authors, John Broome suggested that the problem is 'devastating' for accounts of decision-making under moral uncertainty; the late Derek Parfit described the problem as 'fatal'.

[3] For discussion of decision-making under moral uncertainty in conditions of merely ordinal theories and level-comparability, see Christian Tarsney, 'Moral Uncertainty for Deontologists', *Ethical Theory and Moral Practice*, vol. 21, no. 3 (2018), pp. 505–20. https://doi.org/10.1007/s10677-018-9924-4

provide a concept of choice-worthiness that is at least interval-scale measurable. That is, you need to be able to make sense, on every theory in which you have credence, of the idea that differences in choice-worthiness can be compared—that, for instance, the difference between the choice-worthiness of killing and that of lying is greater than the difference between the choice-worthiness of lying and that of withholding some insignificant truth.

Second, you need to be able to compare the magnitude of the difference in choice-worthiness across different moral theories. That is, you need to be able to tell whether the difference in choice-worthiness between *A* and *B*, on T_i, is greater than, smaller than, or equal to, the difference in choice-worthiness between *C* and *D*, on T_j. Moreover, you need to be able to tell, at least roughly, *how much* greater the choice-worthiness difference between *A* and *B* on T_i is than the choice-worthiness difference between *C* and *D* on T_j.

Many theories do provide interval-scale measurable choice-worthiness: in general, if a theory orders empirically uncertain prospects in terms of their choice-worthiness and the choice-worthiness relation satisfies the axioms of expected utility theory, then the theory provides interval-scale measurable choice-worthiness.[4] Many theories satisfy these axioms. Consider, for example, the version of utilitarianism according to which one should maximize expected wellbeing (and which therefore satisfies the axioms of expected utility theory[5]). If, according to this form of utilitarianism, a guarantee of saving person *A* is equal to a 50% chance of saving no one and a 50% chance of saving both persons *B* and *C*, then we would know that, according to this form of utilitarianism, the difference in choice-worthiness between saving person *B* and *C*, and saving person *A*, is the same as the difference in choice-worthiness between saving person *A* and saving no one. We give meaning to the idea of comparing differences in choice-worthiness by appealing to what the theory says in cases of uncertainty.

However, this method cannot be applied to all theories. Sometimes, the axioms of expected utility theory clash with common-sense intuition, such

[4] As shown in Von Neumann and Morgenstern, *Theory of Games and Economic Behavior*. The application of this idea to moral theories is discussed at length in John Broome, *Weighing Goods: Equality, Uncertainty, and Time*, Cambridge, MA: Basil Blackwell, 1991.

[5] For the purpose of this discussion, we assume away the possibility of infinite amounts of value (which would mean that the view violates the Archimidean axiom). Alternatively, one could replace the view we discuss with one on which moral value is bounded above and below.

as in the Allais paradox.[6] If a theory is designed to cohere closely with common-sense intuition, as many non-consequentialist theories are, then it may violate these axioms. And if the theory does violate these axioms, then, again, we cannot use probabilities in order to make sense of interval-scale measurable choice-worthiness.

Plausibly, Kant's ethical theory is an example of a merely ordinally measurable theory.[7] According to Kant, murder is less choiceworthy than lying, which is less choiceworthy than failing to aid someone in need. But we don't think it makes sense to say, even roughly, that on Kant's view the difference in choice-worthiness between murder and lying is greater than or less than the difference in choice-worthiness between lying and failing to aid someone in need. So someone who has non-zero credence in Kant's ethical theory simply can't use expected choice-worthiness maximization over all theories in which she has credence.

The second problem for the maximizing expected choice-worthiness account is the problem of intertheoretic comparisons. Even when theories do provide interval-scale measurable choice-worthiness, there is no guarantee that we will be able to compare magnitudes of choice-worthiness differences between one theory and another. Previously, we gave the example of comparing the difference in choice-worthiness between killing one person to save fifty and refraining from doing so, according to a rights-based moral theory and according to utilitarianism. In this case, there's no intuitive answer to the question of whether the situation is higher-stakes for the rights-based theory than it is for utilitarianism or vice versa. And in the absence of intuitions about the case, it's difficult to see how there could be any way of determining an answer. We'll discuss this issue more in Chapters 4 and 5.

The question of what to do when we cannot make intratheoretic comparisons of units of choice-worthiness (that is, those theories are merely ordinal), and when we can make neither unit nor level comparisons of choice-worthiness across theories, has not been discussed in the literature. At best, it has been assumed that, in the absence of intertheoretic comparisons, the only alternative to maximizing expected choice-worthiness is

[6] Maurice Allais, 'Allais Paradox', in John Eatwell, Murray Milgate, and Peter Newman (eds), *The New Palgrave: A Dictionary of Economics*, London: Macmillan, 1987, vol. 1, pp. 78–80.

[7] Kant's ethics violates at least the *continuity* assumption: that, for three options A, B, and C, such that A is at least as choiceworthy as B, which is at least as choiceworthy as C, there exists a probability p such that B is equally as choiceworthy as $p \times A + (1 - p) \times C$.

the account according to which one should simply act in accordance with *My Favorite Theory* or *My Favorite Option*.[8] For that reason, it has been assumed that the lack of intertheoretic comparisons would have drastic consequences. For example, because intertheoretic incomparability entails that *maximize expected choice-worthiness* cannot be applied, Jacob Ross says: 'the denial of the possibility of intertheoretic value comparisons would imply that among most of our options there is no basis for rational choice. In other words, it would imply the near impotence of practical reason'.[9] In a similar vein, other commentators have regarded the problem of intertheoretic comparisons as fatal to the very idea of developing a normative account of decision-making under moral uncertainty. In one of the first modern articles to discuss decision-making under moral uncertainty,[10] James Hudson says:

> Hedging will be quite impossible for the ethically uncertain agent . . . Under the circumstances, the two units [of value, according to different theories] must be incomparable by the agent, and so there can be no way for her [moral] uncertainty to be taken into account in a reasonable decision procedure. Clearly this second-order hedging is impossible.[11]

Likewise, Edward Gracely argues, on the basis of intertheoretic incomparability, that:

> the proper approach to uncertainty about the rightness of ethical theories is to determine the one most likely to be right, and to act in accord with its dictates. Trying to weigh the importance attached by rival theories to a particular act is ultimately meaningless and fruitless.[12]

[8] E.g. Ross, 'Rejecting Ethical Deflationism', p. 762, fn.11.

[9] Note that Ross uses this purported impotence as a *reductio* of the idea that different theories' choice-worthiness rankings can be incomparable. However, if our argument in the preceding paragraphs is sound, then Ross's position is not tenable.

[10] The first modern article published on the topic of moral uncertainty appears to be Ted Lockhart, 'Another Moral Standard', *Mind*, vol. 86, no. 344 (October 1977), pp. 582–6, followed by James R. Greenwell, 'Abortion and Moral Safety', *Crítica*, vol. 9, no. 27 (December 1977), pp. 35–48 ad Raymond S. Pfeiffer, 'Abortion Policy and the Argument from Uncertainty', *Social Theory and Practice*, vol. 11, no. 3 (Fall 1985), pp. 371–86. We thank Christian Tarsney for bringing these articles to our attention.

[11] Hudson, 'Subjectivization in Ethics', p. 224.

[12] Gracely, 'On the Noncomparability of Judgments Made by Different Ethical Theories', pp. 331–2.

The above philosophers don't consider the idea that different criteria could apply depending on the informational situation of the agent. It is this assumption that leads to the thought that the problem of intertheoretic comparisons of value is fatal for accounts of decision-making under moral uncertainty. Against Ross and others, we'll argue that decision-making in conditions of moral uncertainty and intertheoretic incomparability is not at all hopeless. In this chapter, we focus on decision-making in conditions of merely ordinal theories. In the next chapter, we focus on decision-making when theories are interval-scale measurable but not comparable.[13] In both cases, we will exploit an analogy between decision-making under moral uncertainty and social choice. So let's turn to that now.

II. Moral Uncertainty and the Social Choice Analogy

Social choice theory, in the 'social welfare functional' framework developed by Amartya Sen,[14] studies how to aggregate individuals' utility functions (where each utility function is a numerical representation of that individual's preferences over social states) into a single 'social' utility function, which represents 'social' preferences over social states, i.e. which state is better than another. A *social welfare functional* is a function from sets of utility functions to a 'social' utility function. Familiar examples of social welfare functionals include: utilitarianism, according to which *A* has higher social utility than *B* iff the sum total of utility over all individuals is greater for *A* than for *B*; and maximin, according to which *A* has higher social utility than *B* iff *A* has more utility than *B* for the worst-off member of society.

Similarly, the theory of decision-making under moral uncertainty studies how to aggregate different theories' choice-worthiness functions into a single appropriateness ordering. The formal analogy between these two disciplines should be clear.[15] Instead of individuals we have theories; instead of

[13] For the purpose of these chapters, we put the issue of *intra*theoretic incomparability to the side, and only consider theories that have complete choice-worthiness orderings.

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

[15] Note that this analogy is importantly different from other analogies between decision theory and social choice theory that have recently been drawn in the literature. Rachael Briggs's analogy ('Decision-Theoretic Paradoxes as Voting Paradoxes', *Philosophical Review*, vol. 119, no. 1 (January 2010), pp. 1–30) is quite different from ours: in her analogy, a decision theory is like a voting theory but where the voters are the decision-maker's future selves. Samir Okasha's analogy ('Theory Choice and Social Choice: Kuhn versus Arrow', *Mind*, vol. 120, no. 477 (January 2011), pp. 83–115) is formally similar to ours, but his analogy is between the problem

Table 3.1

Social Choice Theory	⇒	Moral Uncertainty
Individuals	⇒	First-order moral theories
Individual utility	⇒	Choice-worthiness function
Social welfare functional	⇒	Theory of decision-making under moral uncertainty
Utilitarianism	⇒	Maximize expected choice-worthiness

preferences we have choice-worthiness orderings; and rather than a social welfare functional we have a theory of decision-making under moral uncertainty. And, just as social choice theorists try to work out what the correct social welfare functional is, so we are trying to work out what the correct theory of decision-making under moral uncertainty is. Moreover, just as many social choice theorists tend to be attracted to weighted utilitarianism ('weighted' because the weights assigned to each individual's welfare need not be equal) when information permits,[16] so we are attracted to its analogue under moral uncertainty, *maximize expected choice-worthiness,* when information permits (see Table 3.1).

The formal structure of the two problems is very similar. But the two problems are similar on a more intuitive level as well. The problem of social choice is to find the best compromise in a situation where there are many people with competing preferences. The problem of moral uncertainty is to find the best compromise in a situation where there are many possible moral theories with competing recommendations about what to do.

What's particularly enticing about this analogy is that the literature on social choice theory is well developed, and results from social choice theory might be transferable to moral uncertainty, shedding light on that issue. In particular, since the publication of Amartya Sen's *Collective Choice and Social Welfare*,[17] social choice theory has studied how different social welfare functionals may be axiomatized under different *informational assumptions*. One can vary informational assumptions in one of two ways. First, one can vary the *measurability assumptions*, and, for example, assume that utility is

of social choice and the problem of aggregating different values within a pluralist epistemological theory, rather than the problem of aggregating different values under moral uncertainty.

[16] For the reasons why, given interval-scale measurable and interpersonally comparable utility, weighted utilitarianism is regarded as the most desirable social choice function see, for example, Charles Blackorby, David Donaldson, and John A. Weymark, 'Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction', *International Economic Review*, vol. 25, no. 2 (1984), pp. 327–56.

[17] Sen, *Collective Choice and Social Welfare*.

merely ordinally measurable, or assume that it is interval-scale measurable. Second, one can vary the *comparability assumptions*: one can assume that we can compare differences in utility between options across different individuals; or one can assume that such comparisons are meaningless. The problem of determining how such comparisons are possible is known as the problem of interpersonal comparisons of utility. As should be clear from the discussion in the previous section, exactly the same distinctions can be made for moral theories: choice-worthiness can be ordinally or interval-scale measurable; and it can be intertheoretically comparable or incomparable.

Very roughly, what is called voting theory is social choice theory in the context of preferences that are non-comparable and merely ordinally measurable. Similarly, the problem with which we're concerned in this chapter is how to aggregate individual theories' choice-worthiness functions into a single appropriateness ordering in conditions where choice-worthiness is merely ordinally measurable.[18] So we should explore the idea that voting theory will give us the resources to work out how to take normative uncertainty into account when the decision-maker has non-zero credence only in merely ordinal theories.

However, before we begin, we should note two important disanalogies between voting theory and decision-making under moral uncertainty. First, theories, unlike individuals, don't all count for the same: theories are objects of credences. The answer to this disanalogy is obvious. We treat each theory like an individual, but we weight each theory's choice-worthiness function in proportion with the credence the decision-maker has in that the theory. So the closer analogy is with weighted voting.[19]

The second and more important disanalogy is that, unlike in social choice, a decision-maker under moral uncertainty will face varying information from different theories at one and the same time. For a typical decision-maker under moral uncertainty, some of the theories in which she has credence will be interval-scale measurable and intertheoretically comparable; others will be interval-scale measurable but intertheoretically incomparable; others again will be merely ordinally measurable. In contrast, when social choice theorists study different informational set-ups, they generally assume that the same informational assumptions apply to all individuals.

[18] And, as noted previously, we assume that comparisons of *levels* of choice-worthiness are not possible between theories.

[19] An example of a weighted voting system is the European Council, where the number of votes available to each member state is proportional to that state's population.

We discuss this issue at the end of Chapter 3, providing a general theory of decision-making under moral uncertainty where the precise method of aggregating the decision-maker's uncertainty is sensitive to the information provided by the theories in which she has credence, but which can be applied even in cases of varying informational conditions. In this chapter, however, we assume that all theories in which the decision-maker has credence are merely ordinal. With these caveats, the obvious next question is: which voting system should we use as an analogy?

III. Some Voting Systems

In the previous chapter, we looked at *My Favorite Theory* and *My Favorite Option*. One key argument against them was that they are insensitive to magnitudes of choice-worthiness differences. But if we are considering how to take normative uncertainty into account given that a decision-maker only has non-zero credence in merely ordinal theories, then this objection does not apply. So one might think MFT or MFO gets it right in conditions of merely ordinal theories. However, even in this situation, we think we have good reason to reject these accounts. Consider the following case.[20]

Judge

Julia is a judge who is about to pass a verdict on whether Smith is guilty of murder. She is very confident that Smith is innocent. There is a crowd outside, who are desperate to see Smith convicted. Julia has three options:

A: Pass a verdict of 'guilty'.
B: Call for a retrial.
C: Pass a verdict of 'innocent'.

Julia knows that the crowd will riot if Smith is found innocent, causing mayhem on the streets and the deaths of several people. If she calls for a retrial, she knows that he will be found innocent at a later date, that the crowd will not riot today, and that it is much less likely that the crowd

[20] In the cases that follow, and in general when we are discussing merely ordinal theories, we will refer to a theory's choice-worthiness ordering directly, rather than its choice-worthiness function. We do this in order to make it clear which theories are to be understood as ordinal, and which are to be understood as interval-scale measurable. We use the symbol '>' to mean 'is more choiceworthy than'.

will riot at that later date. If she declares Smith guilty, the crowd will be appeased and go home peacefully. She has credence in three moral theories.

35% credence in a variant of utilitarianism, according to which *A*>*B*>*C*.

34% credence in a variant of common sense morality, according to which *B*>*C*>*A*.

31% credence in a deontological theory, according to which *C*>*B*>*A*.

MFT and MFO both regard *A* as most appropriate, because *A* is both most choiceworthy according to the theory in which the decision-maker has highest credence, and has the greatest probably of being right. But note that Julia thinks *B* is very nearly as likely to be right as is *A*; and she's 100% certain that B is at least second best. It seems highly plausible that this certainty in *B* being at least the second-best option should outweigh the slightly lower probability of *B* being maximally choiceworthy. So it seems, intuitively, that *B* is the most appropriate option: it is well supported in general by the theories in which the decision-maker has credence. But neither MFT nor MFO can take account of that fact. Indeed, MFT and MFO are completely insensitive to how theories rank options that are not maximally choiceworthy. But to be insensitive in this way, it seems, is simply to ignore decision-relevant information. So we should reject these theories.

If we turn to the literature on voting theory, can we do better? Within voting theory, the gold standard voting systems are *Condorcet extensions*.[21] The idea behind such voting systems is that we should think how candidates would perform in a round-robin head-to-head tournament—every candidate is compared against every other candidate in terms of how many voters prefer one candidate to the other. A voting system is a Condorcet extension if it satisfies the following condition: that, if, for every other option *B*, the majority of voters prefer *A* to *B*, then *A* is elected.

We can translate this idea into our moral uncertainty framework as follows. Let's say that *A beats B* (or *B is defeated by A*) iff it is true that, in a

[21] A brief comment on some voting systems we don't consider: we don't consider range voting because we're considering the situation where theories give us only ordinal choice-worthiness, whereas range voting requires interval-scale measurable choice-worthiness. We don't consider instant-runoff (or 'alternative vote') because it violates monotonicity: that is, one can cause *A* to win over *B* by choosing to vote for *B* over *A* rather than vice versa. This is seen to be a devastating flaw within voting theory (see, for example, Nicholas Tideman, *Collective Decisions and Voting*, Routledge (2017)), and we agree: none of the voting systems we consider violate this property.

pairwise comparison between *A* and *B*, the decision-maker thinks it more likely that *A* is more choiceworthy than *B* than that *B* is more choiceworthy than *A*. *A* is the *Condorcet winner* iff *A* beats every other option within the option-set. A theory of decision-making under moral uncertainty is a Condorcet extension if it elects a Condorcet winner whenever one exists. Condorcet extensions get the right answer in *Judge*, because *B* beats both *A* and *C*.

However, often Condorcet winners do not exist. Consider the following 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 that 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*.

In this case, no Condorcet winner exists: *B* beats *C*, *C* beats *A*, but *A* beats *B*. But, intuitively, *C* is more appropriate than *A* or *B*: *A*>*B*>*C*, *B*>*C*>*A*, and *C*>*A*>*B* are just 'rotated' versions of each other, with each option appearing in each position in the ranking exactly once. Given this, then the ranking with the highest credence should win out, and C should be the most appropriate option.

So Condorcet extensions need some way to determine a winner even when no Condorcet winner exists. Let us say that the *magnitude of a defeat* is the difference between the credence the decision-maker has that *A* is more choiceworthy than *B* and the credence the decision-maker has that *B* is more choiceworthy than *A*. A simple but popular Condorcet extension is the Simpson–Kramer method:

Simpson–Kramer Method: *A* is more appropriate than *B* iff *A* has a smaller biggest pairwise defeat than *B*; *A* is equally as appropriate as *B* iff *A* and *B*'s biggest defeats are equal in magnitude.

In *Hiring Decision*, the biggest pairwise defeat for *A* and *B* is 30% to 70%, whereas the biggest pairwise defeat for *C* is only 40% to 60%, so the magnitude of the biggest defeat is 40% for *A* and *B* and only 20% for *C*. So, according to the Simpson–Kramer method, *C* is the most appropriate option, which seems intuitively correct in this case (see Table 3.2).

In what follows, we'll use the Simpson–Kramer Method as a prototypical Condorcet extension.[22] Though Condorcet extensions are the gold standard within voting theory, they are not right for our purposes. Whereas voting systems rarely have to handle an electorate of variable size, theories of decision-making under moral uncertainty do: varying the size of the electorate is analogous to changing one's credences in different moral theories. It's obvious that our credences in different moral theories should often

Table 3.2

	A	*B*	*C*
A		**30%:70%**	70%:30%
B	70%:30%		**40%:60%**
C	**30%:70%**	60%:40%	

[22] There are other Condorcet extensions that are, in our view, better than the Simpson–Kramer method, such as the Schulze method (Markus Schulze, 'A New Monotonic, Clone-Independent, Reversal Symmetric, and Condorcet-Consistent Single-Winner Election Method', *Social Choice and Welfare*, vol. 36, no. 2 (February 2011), pp. 267–303) and Tideman's Ranked Pairs (T. N. Tideman, 'Independence of Clones as a Criterion for Voting Rules', *Social Choice and Welfare*, vol. 4, no. 3 (September 1987), pp. 185–206), because they satisfy some other desirable properties that the Simpson–Kramer method fails to satisfy. However, these are considerably more complex than the Simpson–Kramer Method, and fail to be satisfactory for exactly the same reasons why the Simpson–Kramer method fails to be satisfactory. So in what follows we will just use the Simpson–Kramer method as our example of a Condorcet extension.

change. But Condorcet extensions handle that fact very poorly. A minimal condition of adequacy for handling variable electorates is as follows.[23]

Twin Condition: If an additional voter who has exactly the same preferences as a voter who is already part of the electorate joins the electorate and votes, that does not make the outcome of the vote worse by the lights of the additional voter.

The parallel condition in the case of decision-making under normative uncertainty is:

Updating Consistency: Increasing one's credence in some theory does not make the appropriateness ordering worse by the lights of that theory. More precisely: For all T_i, A, B, if A is more choiceworthy than B on T_i, and A is more appropriate than B, then if the decision-maker increases her credence in T_i, decreasing her credence in all other theories proportionally, it is still true that A is more appropriate than B.

Updating Consistency seems to us to be a necessary condition for any theory of decision-making under moral uncertainty. When all theories in which the decision-maker has non-zero credence are merely ordinally measurable, appropriateness should be determined by two things only: first, how highly ranked the option is, according to the theories in which the decision-maker has non-zero credence; and, second, how much credence the decision-maker has in those theories. It would be perverse, therefore, if increasing one's credence in a particular theory on which *A* is more choiceworthy than *B* makes *A* less appropriate than *B*.

However, all Condorcet extensions violate that condition. To see this, consider the following case.

Tactical Decisions

Jane is a military commander. She needs to take aid to a distant town, through enemy territory. She has four options available to her:

A: Bomb and destroy an enemy hospital in order to distract the enemy troops in the area. This kills 10 enemy civilians. All 100 of her soldiers and all 100 enemy soldiers survive.

[23] First given in Hervé Moulin, 'Condorcet's Principle Implies the No Show Paradox', *Journal of Economic Theory*, vol. 45, no. 1 (June 1988), pp. 53–64.

B: Bomb and destroy an enemy ammunitions factory, restricting the scale of the inevitable skirmish. This kills 10 enemy engineers, who help enemy soldiers, though they are not soldiers themselves. As a result, 90 of her soldiers and 90 enemy soldiers survive.

C: Status quo: don't make any pre-emptive attacks and go through the enemy territory only moderately well-armed. 75 of her soldiers and 75 enemy soldiers survive.

D: Equip her soldiers with much more extensive weaponry and explosives. 95 of her soldiers and none of the enemy soldiers survive.

Jane has credence in five different moral views.

She has 5/16 credence in T_1 (utilitarianism), according to which one should simply minimize the number of deaths. According to T_1, $A>B>C>D$.

She has 3/16 credence in T_2 (partialist consequentialism), according to which one should minimize the number of deaths of home soldiers and enemy civilians and engineers, but that deaths of enemy soldiers don't matter. According to T_2, $D>A>B>C$.

She has 3/16 credence in T_3 (mild non-consequentialism), according to which one should minimize the number of deaths of home soldiers and enemy civilians and engineers, that deaths of enemy soldiers don't matter, and that it's mildly worse to kill someone as a means to an end than it is to let them die in battle. According to T_3, $D>A>C>B$.

She has 4/16 credence in T_4 (moderate non-consequentialism), according to which one should minimize the number of deaths of all parties, but that there is a side-constraint against killing a civilian (but not an engineer or soldier) as a means to an end. According to T_4, $B>C>D>A$.

She has 1/16 credence in T_5 (thoroughgoing non-consequentialism), according to which one should minimize the number of deaths, but that there is a side-constraint against killing enemy civilians or engineers as a means to an end, and that killing enemy civilians as a means to an end is much worse than killing enemy engineers. According to T_5, $C>D>B>A$.

Given her credences, according to the Simpson–Kramer method *D* is the most appropriate option.[24] The above case is highly complicated, and we

[24] *A*'s biggest pairwise defeat is to *D*, losing by 6/16. *B*'s biggest pairwise defeat is to *A*, losing both by 6/16; *C*'s biggest pairwise defeat is to *B*, losing by 8/16; *D*'s biggest pairwise defeat is to *C*, losing by 4/16. So *D* is the most appropriate option according to the Simpson–Kramer method.

have no intuitions about what the most appropriate option is for Jane, so we don't question that answer. However, what's certain is that gaining new evidence in favour of one moral theory, and increasing one's credence in a moral theory, should not have the consequence of making an option which is *worse* by the lights of the theory in which one has increased one's credence *more appropriate*. But that's exactly what happens on the Simpson–Kramer method. Let us suppose that Jane hears new arguments, and increases her credence in T_5 so that now she has 5/20 credence in T_5. The ratios of her credences in all other theories stays the same: she has 5/20 in T_1, 3/20 in T_2, 3/20 in T_3 and 4/20 in T_4. After updating in favour of T_5, *B* becomes the most appropriate option, according to the Simpson–Kramer method.[25] But T_5 regards *D* as more choiceworthy than *B*. So the fact that Jane has updated in favour of T_5 has made the most appropriate option worse by T_5's lights. This is highly undesirable. So we should reject the Simpson–Kramer method.

In fact, it has been shown that *any* Condorcet extension will violate the *Twin Condition* described above;[26] and so any analogous theory of decision-making under moral uncertainty will violate *Updating Consistency*. So, rather than just a reason to reject the Simpson–Kramer method, violation of *Updating Consistency* gives us a reason to reject all Condorcet extensions as theories of decision-making under moral uncertainty.

Before moving on to a voting system that does better in the context of decision-making under moral uncertainty, we'll highlight one additional reason that is often advanced in favour of Condorcet extensions. This is that Condorcet extensions are particularly immune to strategic voting: that is, if a Condorcet extension voting system is used, there are not many situations in which a voter can lie about her preferences in order to bring about a more desirable outcome than if she had been honest about her preferences.

It should be clear that this consideration should bear no weight in the context of decision-making under moral uncertainty. We have no need to worry about theories 'lying' about their choice-worthiness function (whatever that would mean). The decision-maker knows what moral theories she

[25] *A*'s biggest pairwise defeat is to *D*, losing by 10/20. *B*'s biggest pairwise defeats are to *A* and *D*, losing both by 2/20; *C*'s biggest pairwise defeat is to *B*, losing by 5/20; *D*'s biggest pairwise defeat is to *B*, losing by 8/20. *B* has the smallest biggest defeat. So *B* is the most appropriate option according to the Simpson–Kramer method. The Schulze method and Ranked Pairs (mentioned in footnote 20 above) both give exactly the same answers in both versions of *Tactical Decisions*, so this case is a counterexample to them too.

[26] The proof of this is too complex to provide here, but can be found in Moulin, 'Condorcet's Principle Implies the No Show Paradox'.

has credence in, and she knows their choice-worthiness functions. So, unlike in the case of voting, there is no gap between an individual's stated preferences and an individual's true preferences.

IV. The Borda Rule

We have seen that MFT, MFO, and Condorcet extensions do not provide the basis for a plausible theory of decision-making under moral uncertainty. Let's now look at a voting system that does better: the Borda Rule. To see both the Borda Rule's similarity to, and difference from, Condorcet extensions, again we should imagine that all options compete against each other in a round-robin head-to-head tournament. Like the Simpson–Kramer method, the magnitudes of the victories and defeats in these pairwise comparisons matter (where the 'magnitude' of a victory is given by the number of votes in favour of the option minus the number of votes against that option). However, rather than focusing on the size of the biggest pairwise defeat, as the Simpson–Kramer method does, the Borda Rule regards the success of an option as equal to the sum of the magnitudes of its pairwise victories against all other options. The most appropriate option is the option whose sum total of magnitudes of victories is greatest.

To see the difference, imagine a round-robin tennis tournament, with players *A*–*Z*. Player *A* beats all other players, but in every case wins during a tiebreaker in the final set. Player *B* loses by only two points to *A*, but beats all other players in straight sets. Condorcet extensions care first and foremost about whether a player beats everyone else, and would regard Player *A* as the winner of the tournament. The Borda Rule cares about how many points a player wins in total, and would regard Player *B* as the winner of the tournament. It's not obvious to us which of these two approaches is correct when it comes to moral uncertainty: the arguments for choosing Player *A* or Player *B* both have something going for them. But the fact that it's not obvious shows that we shouldn't reject outright all theories of decision-making under moral uncertainty that aren't Condorcet extensions.

Defining the *Borda Rule* more precisely:

An option *A*'s *Borda Score*, for any theory T_i, is equal to the number of options within the option-set that are less choiceworthy than *A* according to theory T_i's choice-worthiness function, minus the number of options within the option-set that are more choiceworthy than *A* according to T_i's choice-worthiness function.[27]

An option *A*'s *Credence-Weighted Borda Score* is the sum, for all theories T_i, of the Borda Score of *A* according to theory T_i multiplied by the credence that the decision-maker has in theory T_i.

These definitions allow us to state the Borda Rule:

Borda Rule: An option *A* is more appropriate than an option *B* iff *A* has a higher Credence-Weighted Borda Score than *B*; *A* is equally as appropriate as *B* iff *A* and *B* have an equal Credence-Weighted Borda Score.

In this way, the Borda Rule generates not just a set of maximally appropriate actions, but also an appropriateness function.

We can argue for the Borda Rule in two ways. First, we can appeal to cases. Consider again the *Judge* case. We criticized MFT and MFO for not being sensitive to the entirety of the decision-maker's credence distribution, and for not being sensitive to the entire range of each theory's choice-worthiness ordering. The Borda Rule does not make the same error. In *Judge*, the Borda Rule ranks *B* as most appropriate, then *C*, then *A*.[28] This seemed to us to be the intuitively correct result: favouring an option that is generally well-supported rather than an option that is most choiceworthy according to one theory but least choiceworthy according to all others. In *Hiring*

[27] The reader might have seen an option's Borda Score defined as equal simply to the number of options below it. The addition of 'minus the number of options that rank higher' clause is the most common way of accounting for tied options. The motivation for this way of dealing with ties is that we want the sum total of Borda Scores over all options to be the same for each theory, whether or not that theory claims there are tied options; if we did not do this, we would be giving some moral theories greater voting power on arbitrary grounds. We will return to whether this account is accurate, suggesting that it should be slightly amended, in the next chapter. The reader may also have seen a Borda Score defined such that an option ranked *i*th receives $n - i$ points plus 0.5 for every option with which it is tied, where *n* is the total number of options in the option-set. This definition is equivalent to ours; however, ours will prove easier to use when it comes to extending the account in the next section.

[28] Because it doesn't affect the ranking when there are no ties, when giving working we will use a simpler definition of a Borda Score: that an option's Borda Score, for some theory T_i, is equal to the number of options below it on T_i's choice-worthiness ranking. Given this definition, option *A* receives a score of $35 \times 2 + 0 + 0 = 70$; option *B* receives a score of $35 + 34 \times 2 + 31 = 134$; option *C* receives a score of $0 + 34 + 31 \times 2 = 96$.

Decision, according to the Borda Rule, C is the most appropriate candidate, then A then B.[29] Again, this seems to us to be obviously the correct answer.

Finally, consider the *Tactical Decisions* case. In this case, according to the Borda Rule, before updating, the most appropriate option for Jane is A, followed by B, then D, then C.[30] As we said before, we don't have any intuitions in this case about which option is most appropriate. But we do know that Jane increasing her credence in T_5 (which ranks $C>D>B>A$) shouldn't make the most appropriate option worse by T_5's lights. Indeed, given that it seems unclear which option is most appropriate, we would expect a substantial increase in her credence in T_5 to improve the appropriateness ranking by T_5's lights. And that's what we find. After updating in favour of T_5, according to the Borda Rule, the appropriateness ranking is D, followed by B, then C, then A.[31]

However, appeal to cases is limited in its value because we can't know whether the cases we have come up with are representative, or whether there exist other cases that are highly damaging to our favoured proposal that we simply haven't thought of. A better method is to appeal to general desirable properties. One such property is *Updating Consistency*. In the context of voting theory, it has been shown that, among the commonly discussed and plausible voting systems, only *scoring rules* satisfy the equivalent property, where a scoring rule is a rule that gives a score to an option based on its position in an individual's preference ranking, and claims you should maximize the sum of that score across individuals.[32] The Borda Rule is an example of a scoring rule, as is MFO, whereas MFT and the Simpson–Kramer Method are not. But we rejected MFO on the grounds that it wasn't sensitive to the entirety of theories' choice-worthiness rankings. So we could add in another intuitively obvious condition that the score of each option in i^{th} position has to be strictly greater than the score given to an option in $(i+1)^{\text{th}}$ position. This wouldn't quite single out the Borda Rule, but it would come close.

[29] Option A receives a score of $30 \times 2 + 0 + 10 \times 1 = 100$. Option B receives a score of $30 \times 1 + 30 \times 2 + 0 = 90$. Option C receives a score of $0 + 30 \times 1 + 40 \times 2 = 110$. So, on the Borda Rule, $A>B>C$.

[30] Option A receives a score of $5 \times 3 + 3 \times 2 + 3 \times 2 + 0 + 0 = 27$. B's score is $5 \times 2 + 3 \times 1 + 0 + 4 \times 3 + 1 \times 1 = 26$. C's score is $5 \times 1 + 0 + 3 \times 1 + 4 \times 2 + 1 \times 3 = 19$. D's score is $0 + 3 \times 3 + 3 \times 3 + 4 \times 1 + 1 \times 2 = 24$.

[31] Option A receives a score of $5 \times 3 + 3 \times 2 + 3 \times 2 + 0 + 0 = 27$. B's score is $5 \times 2 + 3 \times 1 + 0 + 4 \times 3 + 4 \times 1 = 29$. C's score is $5 \times 1 + 0 + 3 \times 1 + 4 \times 2 + 4 \times 3 = 28$. D's score is $0 + 3 \times 3 + 3 \times 3 + 4 \times 1 + 4 \times 2 = 30$.

[32] See Moulin, 'Condorcet's Principle Implies the No Show Paradox'.

In order to fully axiomatize the Borda Rule, we need another condition, as follows.

Cancellation: If, for all pairs of options (A,B), S thinks it equally likely that $A>B$ as that $B>A$, then all options are equally appropriate.[33]

It has been shown that the only scoring function that satisfies the voting system analogue of *Cancellation* is the Borda Rule.[34]

One might question *Cancellation* on the following grounds. Consider a case where one has 50% credence in a theory according to which $A>B>C$, and 50% credence in a theory according to which $C>B>A$. One might think that B is the most appropriate option (even though, according to *Cancellation*, all three options are equally appropriate). The grounds for this might be the ordinal equivalent of risk-aversion, whereas the Borda Rule incorporates the equivalent of risk-neutrality. However, in Chapter 1, we endorsed risk-neutral MEC as a default view. If you should be risk-neutral when you can maximize expected choice-worthiness, then surely you should be risk neutral in the ordinal case as well. So for that reason we suggest that the Borda Rule should be the default theory of decision-making in the face of merely ordinal moral theories.

Conclusion

The problem of intertheoretic comparisons is generally considered to be *the* problem facing normative accounts of decision-making under moral uncertainty. It is often assumed that, if theories are intertheoretically incomparable, then all accounts of decision-making under moral uncertainty are doomed—we should just go back to ignoring moral uncertainty, or to assuming our favorite moral theory to be true when deciding what to do.

This chapter has shown the above assumption to be false. How to act in light of moral uncertainty should be sensitive to the information that theories give the decision-maker. And even in the situation in which choice-worthiness is merely ordinally measurable across all theories in

[33] The voting system analogue is: if for all pairs of alternative (x,y), the number of voters preferring x to y equals the number of voters preferring y to x, then a tie between all options should be declared. See H. P. Young, 'An Axiomatization of Borda's Rule', *Journal of Economic Theory*, vol. 9, no. 1 (September 1974), pp. 43–52.

[34] Young, 'An Axiomatization of Borda's Rule'.

which the decision-maker has non-zero credence, there is a plausible way to take decision-theoretic uncertainty into account, namely the Borda Rule.

However, even in conditions of intertheoretic incomparability we often have more information than merely ordinal information. Theories can give interval-scale measurable choice-worthiness, yet be incomparable with each other. How to take moral uncertainty into account in that informational condition is the subject of the next chapter.