

Abstract

Moral Uncertainty and Bargaining

Harry Reginald Lloyd

2025

I develop and defend an original answer to the problem of moral uncertainty – the question of how to behave when one has at least some credence in more than one moral theory. My *marketplace* response to this problem models the moral theories in which one has credence using ‘theory representatives,’ who bargain with each other to determine how the morally uncertain decision maker is to behave. This marketplace approach gives plausible recommendations in a wide range of cases, and also escapes the various objections that have been levelled against other proposed solutions to the problem of moral uncertainty.

Moral Uncertainty and Bargaining

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Harry Reginald Lloyd

Dissertation Director: Shelly Kagan

December 2025

© 2026 by Harry Reginald Lloyd

All rights reserved.

Contents

1	Introduction	1
1.1	Moral uncertainty	1
1.2	MEC versus IMM	4
1.3	Motivating IMM	9
2	Resource division	21
2.1	Initial endowments	22
2.2	Proportional division	27
2.3	Alternatives	32
2.4	Contracts	42
2.5	Trade	51
3	Risk	67
4	Intertemporal dynamics	73
4.1	Intertemporal bargaining	73
4.2	Noncompliance	79

4.2.1	Responses	81
4.2.2	Compensation	90
4.2.3	Individual shortfall	95
4.2.4	Overall shortfall	113
4.3	Descriptive updating	118
4.4	Moral updating	131
5	Discrete choice	141
5.1	Divisible sequences	143
5.2	Lotteries	153
5.3	Problems with lotteries	159
5.3.1	Stochasticity	159
5.3.2	Risk attitudes	162
5.4	Social planning	166
5.4.1	Augmenting IMM	166
5.4.2	The sequence criterion	172
5.4.3	Simple compensation	177
5.4.4	Compensatory scope	183
5.4.5	Relative stakes	188
5.4.6	Discrete-choice sets	194
5.4.7	Sequences	199
5.5	Complications	202
5.5.1	Repetition	202

CONTENTS

v

5.5.2	Division	207
5.6	Synthesis	209
6	Prerogatives	217
7	Bargaining	229
7.1	Cooperative solutions	229
7.1.1	Nash bargaining	230
7.1.2	Disagreement utilities	232
7.1.3	Scale invariance	237
7.1.4	Transformations	239
7.1.5	Interval scales	244
7.2	Subvaluation	251
7.3	Shortfalls	264
8	Conclusion	271
8.1	Evaluation	271
8.2	Future research	275
8.3	Coda	281
	List of choice situations	283
	References	291

Dedication

For Alan and Kathleen

*I like best of all, in New York, the granite rocks in Central Park,
as beautiful in their rugged rhythm as any to be found on high
mountaintops. . . None there are, I suspect, who share with me the
delight in the rocks – those silent, immutable rocks.*

– Lin Yutang, *With Love and Irony*

Acknowledgements

For helpful comments and conversations, I wish to thank Stephen Darwall, Conor Downey, Paul Forrester, Hilary Greaves, Daniel Greco, Patrick Kaczmarek, Shelly Kagan, Andreas Mogensen, Marcus Pivato, Michael Plant, Stefan Riedener, Larry Samuelson, Sun-Joo Shin, Christian Tarsney, and Martin Vaeth. I also wish to thank the Forethought Foundation and the Happier Lives Institute for their financial support.

Chapter 1

Introduction

1.1 Moral uncertainty

Almost all of us find ourselves uncertain about at least some aspects of morality. Can saving someone's pride ever be sufficient moral reason to lie? Is vegetarianism morally required? And am I morally permitted to prioritize my friends and family over the needs of distant strangers? Even many professional ethicists are uncertain about normative questions like these. This means that we often have to make important decisions whilst being in the dark about morality requires of us.

Still, even if I am highly uncertain about the rightness or wrongness of the various options available to me, I might nonetheless be able to work out how it is most *appropriate* for me to behave in light of my moral uncertainty.¹ For

¹'Appropriate' here is a term of art, that different philosophers explicate in different ways. Perhaps the most popular approach is to define appropriateness in terms of

instance, suppose that I am uncertain about which moral theory is correct, and that I need to decide which charity to donate \$100 to. Suppose I have 60% credence in a moral theory according to which it is quite important for me to donate all of my money to Oxfam, and 40% credence in a moral theory according to which it is morally unimportant which charity I donate to.² Under these conditions of moral uncertainty, there is plausibly some sense in which the only ‘appropriate’ option – taking into account the views of both of the moral theories in which I have credence – is for me to donate to Oxfam.

Of course, seeing what it is appropriate to do in this particular, simplified case does not yet give one any guidance about how to behave appropriately in other, more complicated cases. What we would really like to have is a general criterion of appropriate choice under conditions of moral uncertainty. That’s what this dissertation aims to provide.

At first, it might be quite natural to suppose that we should handle moral uncertainty analogously to how we should handle descriptive uncertainty. Rational decision making under descriptive uncertainty about the consequences

rationality (e.g. Sepielli 2009; 2014; Bykvist 2014; Geyer 2018; MacAskill, Bykvist and Ord 2020, pp. 20-1; Tarsney 2024). Alternatively, Pittard and Worsnip (2017) defend a metanormative contextualist analysis of appropriateness, according to which the ‘ought’ of appropriateness is on all fours with the ‘ought’s of objective and subjective first-order morality. And Olle Risberg (2023) argues that whereas first-order moral theories are truth-apt answers to “the question of what we *ought* to do,” theories of appropriateness are conative responses to “the question of what *to* do.” None of my discussion in this dissertation will assume any particular metaethic of appropriateness.

²For some empirical evidence to support the “Bayesian approach to the study of moral judgements,” see Cohen, Nissan-Rozen and Maril 2024.

of one's choices has already been studied extensively by decision theorists, many of whom argue that one should always *maximise expected utility* (I'll give a definition of this view in §1.2 below).³ Inspired by this approach to descriptive uncertainty, many philosophers writing on moral uncertainty have argued that the appropriate response to moral uncertainty is to *Maximise Expected Choiceworthiness* (henceforth: 'MEC')⁴ – a view that I'll discuss in §1.2 below.

However, my main aim in this dissertation will be to develop and defend an alternative theory of appropriateness, which I will refer to as the *Intrapersonal Moral Marketplace* (henceforth: 'IMM') approach. I will motivate this new IMM approach by suggesting that the problem of decision making under moral uncertainty is most closely analogous *not* to the problem of decision making under descriptive uncertainty, but rather to the social problem of finding compromises between different agents with incompatible desires. I will defend this IMM approach by arguing that it has more plausible implications than its rivals like MEC in a range of scenarios.

Before all of that, however, I will begin with a brief discussion of MEC: IMM's chief rival, and the philosophical foil for my presentation of IMM in this dissertation.

³Steele and Stefánsson 2020; Briggs 2023.

⁴Oddie 1994; Lockhart 2000; Sepielli 2009; 2010; Wedgwood 2013; 2017; MacAskill, Bykvist and Ord 2020; MacAskill and Ord 2020; Riedener 2021.

1.2 MEC versus IMM

MEC is perhaps the most popular extant criterion of appropriateness. According to MEC:

(MEC) some option A is appropriate under conditions of moral uncertainty iff choosing A maximises intertheoretic expected choiceworthiness.

The *choiceworthiness* of some option A according to the moral theory T is the strength of the decision maker's all-things-considered moral reason in favour of choosing A according to T.⁵ The *intertheoretic expected choiceworthiness* of each option is a weighted average of its choiceworthinesses according to each of the theories in which the decision maker has credence, where each theory's weight is the decision maker's credence in that theory.⁶

MEC says that we should handle moral uncertainty analogously to how standard decision theory says that we should handle descriptive uncertainty. Suppose that I am at the poker table, and I am uncertain about which cards my opponent has in her hand. According to standard decision theory, I should respond to this descriptive uncertainty by choosing the option (in this case: the bet size) that will maximise expected utility, where the *utility*

⁵MacAskill, Bykvist and Ord 2020, p. 4.

⁶In other words, the intertheoretic expected choiceworthiness of each option can be calculated by multiplying each possible choiceworthiness value for this option by the decision maker's credence that the option has that choiceworthiness value, and then summing over all of these products.

of each possible outcomes measures how much I would value an increase in the probability of that outcome, and where the *expected* utility of each option A is a weighted average of the utilities of the outcomes that might result from choosing option A, where each outcome's weight in this average is my credence that this outcome will eventuate if I in fact choose the option A.

Thus, standard decision theory is designed to select gambles that have the highest possible payoffs in expectation. At the poker table, maximising expected utility might sometimes require me to take a 'safe' bet that I am almost certain will result in an outcome with only modest utility (given my credences about which cards my opponents hold). Yet on other occasions, maximising expected utility might require me to make a 'risky' bet that has a large chance of producing an outcome with negative utility, but also a small chance of producing an outcome with very high positive utility.

Analogously, MEC conceives of the problem of decision making under moral uncertainty as a problem of selecting optimal *gambles* that have the highest possible choiceworthiness in expectation. Sometimes MEC might require choosing a 'safe' option that I am almost certain is only modestly choiceworthy (given my credence distribution over moral theories). But on other occasions, MEC might require choosing a 'risky' option that I think is quite likely to be rather unchoiceworthy, but also has some small likelihood of being very highly choiceworthy. In all of these scenarios, MEC directs the morally uncertain agent to choose an option that optimally trades off risks

against the prospective rewards that she might gamble on achieving.

Many advocates of MEC regard the analogy between it and standard decision theory as a reason to endorse MEC.⁷ For instance, MacAskill, Bykvist and Ord claim that since “expected utility theory is the standard account of how to handle empirical uncertainty . . . maximizing expected choiceworthiness should be the standard account of how to handle moral uncertainty.”⁸ Similarly, Christian Tarsney suggests that treating descriptive and normative uncertainty “differently when we are not forced to is at least *prima facie* inelegant and undermotivated.”⁹

One important disanalogy between descriptive and moral uncertainty concerns intertheoretic choiceworthiness comparisons (this disanalogy troubles many of the philosophers who work on moral uncertainty).¹⁰ If we want to calculate expected utility, then we have to assume that we can compare differences in utility values across rival descriptive hypotheses. In other words: we have to assume that we can compare the values of the different possible outcomes that could eventuate if one or another of our descriptive hypotheses turned out to be correct. Analogously, if we want to calculate expected choiceworthiness, then we have to assume that we can compare differences in choiceworthiness value across rival moral hypotheses or theories. In other

⁷MacAskill, Bykvist and Ord 2020, pp. 47-8; MacAskill and Ord 2020, §6; Tarsney 2021, p. 172; Sepielli 2010, pp. 75-8; although cf. Kaczmarek, Lloyd and Plant 2025.

⁸MacAskill, Bykvist and Ord 2020, pp. 47-8.

⁹Tarsney 2021, p. 172.

¹⁰*Inter alia* Hudson 1989; Gracely 1996; Broome 2012, pp. 184-5; Gustafsson and Torpman 2014; Nissan-Rozen 2015; Hedden 2016, §5.2.1; Newberry and Ord 2021; Gustafsson 2022, §5; Kaczmarek, Lloyd and Plant 2025.

words: we have to assume that we can compare the different possible values that any given moral choice might have if one or another of our moral hypotheses turned out to be correct. However, whilst we clearly find it intelligible to compare utility values across different descriptive hypotheses, it is not so clear that we find it intelligible to compare choiceworthiness values across different moral hypotheses. When we toggle descriptive hypotheses in order to calculate expected utility, we are still holding constant the value theory being used to evaluate the various possible outcomes, and so any value comparisons across our hypotheses are internal to that given value theory, and hence unproblematic. By contrast, if we toggle moral hypotheses to try to calculate expected choiceworthiness, we are *eo ipso* toggling the very value theory being used to evaluate possible choices, and so any value comparisons across these hypotheses will be between different value theories. It is at least an open question whether these value comparisons would even be possible, since it is at least an open question whether for every pair of moral theories there exists some uniquely correct ‘exchange rate’ between those two theories scales for measuring choiceworthiness differences.

MEC also has *prima facie* implausible implications in certain choice situations. For instance, MEC is *fanatical*, in the sense that it can recommend choosing an option that one is almost certain is quite unchoiceworthy, provided this option also has at least some miniscule yet nonzero probability of being extremely choiceworthy.¹¹ Secondly, in certain other kinds of choice

¹¹Baker 2024.

situations, MEC will recommend *prima facie* implausible ‘winner takes all’ resolutions. For example, a morally uncertain philanthropist who has significant credence in several different moral theories without any clear favourite could nonetheless be directed by MEC to donate all of her money to a single charity favoured by only one of the moral theories in which she has credence. Yet many of us intuit that in light of her moral uncertainty, it might well be more appropriate for our philanthropist to *split* her donations between several different charities which have differing ethical priorities.¹² I will discuss this objection to MEC in §2.3 below.

My main aim in this dissertation, though, will not be to litigate the advantages and disadvantages of MEC. Rather, I aim to develop and defend the alternative IMM approach to moral uncertainty, which happens to straightforwardly and elegantly avoid all three of the problems that confront MEC (as I will demonstrate in this dissertation).¹³ Unlike MEC, the IMM approach will not frame the problem of decision making under moral uncertainty as a problem of finding optimal gambles. Rather, the IMM approach will frame the problem as one of finding reasonable *compromises* between the conflicting directives issued by the moral theories in which one has positive credence.¹⁴ In particular, IMM will draw an analogy between (1) the in-

¹²Greaves and Cotton-Barratt (2024, p. 153) report that “anecdotally, as a matter of empirical fact, many people faced with [philanthropic resource division] decision situations . . . feel a powerful pull towards splitting their philanthropic pot”; likewise Karnofsky 2016; 2018; Kaczmarek, Lloyd and Plant 2025.

¹³IMM is not to be confused with the somewhat different *Moral Marketplace Theory* (MMT) developed in Kaczmarek, Lloyd and Plant 2025 (which was an early ancestor of IMM).

¹⁴Even some advocates of MEC have shown some sympathy with this compromise

trapersonal problem of decision making under moral uncertainty, and (2) the interpersonal problems of scarcity and incompatible desires that we confront in our economic interactions with each other.

Over the course of developing IMM in this dissertation, I will highlight several places where we face important theoretical ‘choice points.’ I won’t try to conclusively settle all of these choices here. Rather, my aim will be to give a ‘proof of concept’ that the IMM approach can yield an attractive theory of appropriateness. This aim will be compatible with leaving some of the details unsettled.

1.3 Motivating IMM

Speaking metaphorically, being morally uncertain is a bit like being pulled in several different directions by several different parts of oneself. Speaking personally, perhaps the dominant part of me is utilitarian, and so pulls me in the direction of maximising total welfare. However, perhaps another part of me is Kantian, and so pulls me in the direction of acting only according to maxims that I can will to be universal laws. Each of these parts of me can be thought of as having its own preferences over my behaviour. At least

framing. For instance, in drawing an analogy between moral uncertainty and social choice, Will MacAskill (2016, p. 977) suggests that

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 normative theories with competing recommendations about what to do.

in my own experience, this way of describing moral uncertainty rings true to its distinctive phenomenology. The challenge for a theory of appropriateness is to derive a coherent decision rule from the competing preferences of the different parts of oneself.

This framing of the problem of moral uncertainty suggests an analogy between that problem and those problems of interpersonal decision making that involve multiple agents in some social setting. In particular, I suggest that the problem of decision making under moral uncertainty is closely analogous to the problem of incompatible desires that we confront in our *economic* interactions with each other. In making the case for this analogy, I will begin by characterizing the economic problem in some detail, before turning to the analogy with moral uncertainty.

We can think about our economy as being composed of multiple different households, with each of these households having a certain set of desires.¹⁵ For instance, my household might desire to move to a new house in the country, to take an extra vacation this year, and much else besides. Thinking in terms of social choice, each of these desires has at least some *prima facie* pull in favour of being satisfied, just in virtue of the fact that each of these desires is had by some household or other. If we can better satisfy one household's desires without harming any other household, then we have at

¹⁵I realize that it is slightly strange to describe households as the bearers of desires and preferences (as opposed to the individual people who make up those households). However, microeconomists have traditionally adopted this idiom, and – for reasons that will soon become clear – it suits my purposes to do so too.

least some *prima facie* reason to do this. Moreover, the *pro tanto* force of this reason plausibly co-varies at least with the size of the household. All else being equal, giving a household something that everyone in it desires is more urgent if there are six members than it is if there are only two.

Unfortunately, different households often have desires that pull in different directions; satisfying one household's desires often requires frustrating the desires of other households. We can call this the *social problem of incompatible desires*. One reason for this problem is scarcity: there are too few resources available for every household to be able to satisfy all of their desires. For example, in order to satisfy my household's desire to own a Stradivarius violin, one has to frustrate the desires of other households who would also like to own a Stradivarius, because there are not enough Stradivariuses to go around.

Scarcity is not the only contributor to the social problem of incompatible desires; the problem is also partly caused by logical incompatibilities between desires. Suppose, for instance, that I know some embarrassing things about your personal life. I want to gossip to others about them, but you would prefer for me to keep quiet. Clearly, it is logically impossible for these two desires to be jointly satisfied.

In all of these respects, the intrapersonal problem of moral uncertainty is closely analogous to the social problem of incompatible desires. A morally uncertain agent has credence over multiple different moral theories, with each

of these moral theories directing her to do different things.¹⁶ For instance, a Kantian moral theory might direct her to treat other people as ends, develop her talents, and much else besides. On the other hand, a consequentialist moral theory might direct her to become a vegetarian, donate more money to Oxfam, and much else besides. As the agent sees it, each of these directives plausibly has at least some *prima facie* pull in favour of being followed, just in virtue of the fact that each of these directives is issued by some moral theory in which the agent has positive credence: to the extent that the agent is inclined to *believe* some moral theory T, it seems reasonable for her to be inclined to this extent towards being *guided by* T's directives. Moreover, the *pro tanto* force of this impetus plausibly co-varies at least with the agent's credence in T. All else being equal, following the directives of some moral theory T is more urgent if one has 60% credence in T than it is if one has only 20% credence.

Unfortunately, different moral theories often issue directives that pull in different directions: following one moral theory's directives often requires disobeying the directives of other moral theories. One reason for this is the problem of scarcity.¹⁷ For instance, consider the following choice situation:

¹⁶Whenever I talk about a 'moral theory' in this dissertation, I mean to be referring to something like a maximal consistent set of purely moral propositions. Following common practical in the literature on moral uncertainty, I will assume for the sake of simplicity that our morally uncertain decision makers are epistemically idealised enough to have well-defined credences over 'moral theories' in this sense. However, one important problem swept under the rug by this idealization is the problem of how to handle metaethical noncognitivism. I will discuss this problem in future work.

¹⁷Compare Lockhart 2000, chapter 8.

Philanthropy: some philanthropist is deciding where to donate her fortune. She faces a choice between two charities: one that provides deworming pills to distant children, and a second that supports local soup kitchens. Suppose that this philanthropist has 60% credence in the moral theory T_1 , according to which she should donate her fortune to deworming, and 40% credence in the moral theory T_2 , according to which she should donate her fortune to soup kitchens.

Since the decision maker in **Philanthropy** only has a finite fortune available to donate, her moral uncertainty puts her in a bind. However, if this decision maker instead had unlimited resources, then she could afford to donate unlimited amounts of money to both deworming and soup kitchens – enough for both charities to have more money than they would ever know what to do with. Without resource scarcity, moral uncertainty in **Philanthropy** is unproblematic.

As another example, suppose that some transplant surgeon has positive credence in both utilitarianism and deontology. This surgeon has the option to murder one of her patients in order to harvest and transplant the patient's organs, thereby saving five other people. If there is a shortage of organs for transplant, then the surgeon faces a problem of decision making under moral uncertainty. However, if the surgeon already has unlimited access to organs for transplant, then utilitarianism and deontology will agree that she should not murder her patient. So in this case as well, moral uncertainty is

problematic only when combined with resource scarcity.

However, scarcity is not the only contributor to the problem of decision making under moral uncertainty; the problem is also partly caused by logical incompatibilities between directives. Suppose, for instance, that I can either tell you an uncomfortable truth or a comforting lie. One moral theory in which I have credence says that I am morally required to be honest, but another moral theory in which I have credence says that I am morally required to lie. Clearly, it is logically impossible for these two directives to be jointly obeyed.

Hence, the intrapersonal problem of moral uncertainty is closely analogous to the interpersonal problem of incompatible desires. Thus, perhaps unsurprisingly, the desiderata on solutions to these two problems also strike me as being closely analogous. What are the desiderata on a resolution of the interpersonal problem of incompatible desires? It is plausible to think that an appropriate profile of decisions for a society that is confronted by this problem should be selected by a process that gives each household in that society its due influence over the decision profile. The level of 'due influence' to which each household is entitled should depend at least upon the size of the household, and, to the extent that this is possible, each household should have some of its desires at least partially satisfied in the overall outcome.

Analogously, I suggest that the appropriate life plan for a person with a life like mine and my credence distribution over moral theories should be selected by a process that gives each moral theory in which I have positive

credence its due influence over that life plan. The ‘due influence’ to which each moral theory is entitled should depend at least upon my credence in that moral theory, and, to the extent that this is possible, each moral theory in which I have credence should have at least some of its directives at least partially fulfilled.

Since the desiderata on solutions to the interpersonal problem of incompatible desires are closely analogous to the desiderata on solutions to the intrapersonal problem of moral uncertainty, it is natural to look for a solution to the latter problem by first seeing what we have to come to recognise as an attractive solution to the former problem.¹⁸

At least in theory, under certain idealised modelling assumptions one particularly elegant way of resolving the social problem of incompatible desires in a way that satisfies all of our desiderata is afforded by the twin social technologies of property rights and market exchange. First of all, an initial distribution of property rights determines who is initially entitled to each unit of resources. In principle if not in practice, we can imagine this initial distribution being completely determined by some social planner whose objective is to give each household its ‘fair share’ of property rights over these scarce resources. Each household’s fair share of resources should depend at least in part upon the size of the household. (All else being equal, a six-person household is entitled to a larger initial endowment than a two-person household.) Second, once their initial endowments of property rights have

¹⁸Cf. Plato, *Republic*.

been fixed, our households can then bargain with each other in the market, taking advantage of any opportunities for Pareto improving trades left on the table by the initial distribution of property rights.

As well as handling the problem of resource scarcity, a system of property rights plus market exchange can also handle the problem of logically incompatible desires – at least given a sufficiently broad interpretation of ‘property rights’. Consider, for instance, the case in which I know some embarrassing things about your personal life, and want to gossip to others about them. Supposing that the law does not classify my gossip as defamatory, then I am thereby initially endowed with the right to share this gossip with whomsoever I choose. However, I also have the option to sell you that right, in the form of a nondisclosure agreement (you pay me to sign a contract promising not to share my gossip with anyone). If you value my silence more than I value my liberty to gossip, then at the right price it will be in both of our interests for me to sign the nondisclosure agreement. But if you value my silence less than I value my liberty, then at no price will signing the nondisclosure agreement be in both of our interests, and so I will remain free to gossip.

My IMM approach to moral uncertainty is inspired by this response to the social problem of incompatible desires. Speaking loosely but intuitively, the IMM approach involves giving an initial endowment to each moral theory, and allowing the theories to bargain with each other. More precisely, according to IMM, the appropriate course of action in any given choice situation is determined by a certain *bargaining theoretic model* for the morally uncer-

tain decision maker's choice process in that situation. Each moral theory in which the decision maker has credence is assigned a *theory representative* (or 'representative' for short), and we model the decision maker's behaviour as being determined by a collection of instructions issued by these representatives after they have bargained with each other. Each representative is certain in the truth of the moral theory that they represent, and wants the morally uncertain decision maker to follow the directives of that moral theory as closely as possible.

Each representative is *prima facie* entitled to determine how the decision maker will use a certain share of her total resource endowment. In other words: each representative is initially entitled to the *control rights* over a certain share of the decision maker's resources. In this context, we can understand 'resources' quite broadly, for instance to include things like the decision maker's *time* (as well as, say, her money).

In §2.1 below, I discuss how the initial distribution of control rights over resources across the theory representatives should be determined. Certainly, however, each representative's initial endowment of control rights should depend at least in part upon the decision maker's level of credence in the moral theory that the representative corresponds to. (All else being equal, the representative of a theory in which one has 60% credence should have a larger initial endowment than the representative of a theory in which one has only 20% credence.)

Once their initial endowments of control rights have been fixed, our repre-

sentatives can then bargain with each other in the market, taking advantage of any opportunities for Pareto improving trades left on the table by the initial distribution of control rights. For instance, theory representative R_1 might transfer to theory representative R_2 the right to decide how I will use \$10 worth of my wages this month, in return for R_2 transferring to R_1 the right to decide how I will use one hour of my free time this afternoon. I discuss a case like this in more detail in §2.5 below.

Finally, similarly to how the proprietarian approach to economic life can handle logically incompatible desires just as well as it can handle problems of resource scarcity, IMM's model of control rights plus bargaining can likewise handle the problem of logical incompatibilities between different moral theories' directives. However, this problem of logical incompatibilities introduces a few complications that it will be easiest to discuss after I have first presented IMM's response to resource division problems in more detail. For that reason, I defer discussion of cases involving logical incompatibilities between different moral theories' directives until §5 below.

In this section, I have argued that the problem of decision making under moral uncertainty is closely analogous to the problem of incompatible desires that we confront in our economic interactions with each other. One particularly elegant solution to this economic problem is a system of property rights plus market exchange. Under certain idealized conditions, this system has several attractive properties: the initial distribution of property rights can be arranged such that each household receives its fair share; and then free

trade can realise Pareto improvements left on the table by this initial distribution. This motivates the suggestion that an analogous IMM approach to the problem of moral uncertainty is worthy of investigation.¹⁹ I discuss the details of IMM in the remainder of this dissertation; and I argue that IMM has attractive implications in a range of cases that confound rival views like MEC. I begin by discussing choice situations in which the decision maker has to allocate a continuously divisible resource endowment, including the **Philanthropy** scenario.

¹⁹In at least one respect, market exchange in fact seem even more attractive in the context of IMM's response to the problem of moral uncertainty than they seem in the context of the social problem of incompatible desires. Under idealised conditions of perfect information and perfect rationality, voluntary trades will always be Pareto improving with respect to the agents who agree to the trade: each agent will value what she is buying at least as much as she values what she is selling. Note, however, that this is compatible with some agents benefiting more than others from trade in the market. (For instance, some agents might be lucky in having more potential partners willing to trade with them.) Hence, some egalitarian theorists of distributive justice might object to free trade, because even if we assume an egalitarian distribution of property rights before trade, the outcome after trade might leave some agents comparatively better off than others, and some egalitarians might regard this as 'unfair' to the worse-off agents (cf. Nozick 1974, pp. 160-4, on "how liberty disrupts patterns").

However, these kinds of egalitarian objections to free trade implicitly rely on it being at least in principle possible for us to say that under certain circumstances, some agents are 'comparative better off' than others. (For instance, it might be possible to make comparisons of well-being levels across different individuals.) Thus, in order to object on 'fairness' grounds to the free trade element of IMM's response to the problem of moral uncertainty, one would have to first assume that it is at least in principle possible for us to say that in certain circumstances, some moral theories are 'comparatively better off' than others. (For instance, one might assume that it is possible to make comparisons of choiceworthiness levels across different moral theories.) However, intertheoretic comparability of choiceworthiness strikes many of us as considerably less plausible than, say, interpersonal comparability of well-being. For instance, whereas interpersonal comparisons of hedonic well-being might be grounded in something like physiological comparisons of c-fibres and dopamine, it is much more difficult to think of anything that could ground intertheoretic comparisons of choiceworthiness. Thus, fairness objections to free trade make less sense in the context of IMM's response to moral uncertainty than they make in the context of the social problem of incompatible desires.

Chapter 2

Resource division

Recall that in the **Philanthropy** choice situation (introduced in §1.3), some philanthropist needs to decide how to distribute her fortune between two charities. One of these two charities provides deworming pills to distant children, whereas the other charity supports local soup kitchens. Recall that our philanthropist has 60% credence in the moral theory T_1 , according to which she should donate her fortune to deworming, and 40% credence in the moral theory T_2 , according to which she should donate her fortune to soup kitchens.

In §§2.1-2.2 below, I will explain and defend the IMM response to this simple choice situation. For the sake of simplicity, I will assume throughout this discussion that the choiceworthiness *returns to scale* from funding deworming and soup kitchens are *constant* according to both T_1 and T_2 . In other words, I will assume that according to both of these two moral the-

ories, the strength of the all-things-considered moral reason for or against donating an extra dollar to either of our two charities must always be the same regardless of what our philanthropist does with the rest of her money. It will clearer at the end of §2.2 below why I am making this assumption.

2.1 Initial endowments

In IMM's bargaining model for **Philanthropy**, the two moral theories T_1 and T_2 will be represented by two theory representatives – call them ' R_1 ' and ' R_2 ' – who will bargain with each other to determine how the decision maker should behave. IMM says that when the decision maker first confronts the **Philanthropy** choice situation, R_1 should be initially endowed with control rights over a certain portion of the decision maker's total fortune, with R_2 being endowed with control rights over the remainder.

What should determine the fraction of the decision maker's total fortune initially controlled by R_1 rather than R_2 ? As I noted in §1.3 above, this share should plausibly be increasing in the decision maker's credence in the moral theory T_1 that R_1 represents. Should any other factors also be taken in account?

Suppose for the sake of argument that it is possible to make intertheoretic unit comparisons of choiceworthiness between T_1 and T_2 . In that case, one might think that the initial distribution of control rights should also depend upon how the 'amount of choiceworthiness at stake' in **Philanthropy**

Choiceworthiness	Deworming	Soup kitchens
T_1 : 0.6 credence	5	1
T_2 : 0.4 credence	1	500

Figure 2.1: Potential unit comparisons in **Philanthropy**

according to T_1 compares against the amount of choiceworthiness at stake according to T_2 . For instance, suppose that the T_1 and T_2 's choiceworthiness evaluations can both be measured on a single common scale of choiceworthiness. Furthermore, suppose that the choiceworthiness values of donating an extra dollar to deworming and to soup kitchens according to each of T_1 and T_2 – as measured on this common scale, and regardless of the antecedent distribution of donations to which one might add this extra dollar – are given in figure 2.1.

Under these assumptions, the amount of choiceworthiness at stake in **Philanthropy** is much greater according to T_2 than it is according to T_1 . According to T_2 , changing where any given dollar is donated changes the overall choiceworthiness of one's decision by 499 units; whereas according to T_1 , changing where any given dollar is donated changes the overall choiceworthiness of one's decision by only 4 units. It seems natural enough to think that this fact should *pro tanto* count in favour of R_2 having control over a larger share of the decision maker's total fortune than R_2 would have had otherwise.

Of course, initial endowments of control rights can be sensitive to com-

parisons of the amount of choiceworthiness at stake across different theories in any given choice situation only if it is possible to commensurate between the units of those different theories' respective choiceworthiness scales. As I noted in §1.2, whether we can make these kinds of intertheoretic comparisons is a controversial matter.

As it happens, I myself think that any such comparisons are impossible. So of course, anyone sympathetic to IMM who agrees with me about this point will thereby be committed to rejecting the suggestion that initial endowments of control rights should be sensitive to intertheoretic comparisons between the amounts of choiceworthiness at stake according to the different moral theories in which one has positive credence. On the other hand, it is important to realize that nothing in the IMM approach *per se* rules out believers in intertheoretic comparability adopting a more complicated endowment rule, that is sensitive to both credences and intertheoretic choiceworthiness stakes comparisons.

That being said, in the remainder of this dissertation, I will nonetheless assume that the initial distribution of control rights in any given choice situation should *not* be sensitive to any purported intertheoretic choiceworthiness stakes comparisons. I adopt this assumption for three reasons. (1) As I've already mentioned, I myself regard intertheoretic choiceworthiness comparisons as impossible. But in any case, (2) making this assumption will greatly simplify the remainder of my discussion of IMM. And also (3) many of my claims about IMM in the remainder of this dissertation will not in fact

depend upon how we resolve this theoretical choice point.

Given this assumption that IMM's distribution of control rights across representatives should not be sensitive to inter-theoretic, intra-scenario comparisons of choiceworthiness values, there is perhaps some sense in which IMM is 'insensitive to stakes.' However, as I will discuss in §2.2 below, IMM's verdicts about appropriateness *are* sensitive to *intra-theoretic, inter-scenario* choiceworthiness comparisons, and for that reason there is another reading of 'stakes' on which IMM *is* sometimes sensitive to intertheoretic comparisons of how high the stakes are in various different choice situations.¹

So far in this subsection, I have noted that the initial distribution of control rights to theory representatives should depend upon the decision maker's credence distribution over the corresponding moral theories, and I have adopted the assumption that this initial distribution should not depend upon any purported comparisons between the amounts of choiceworthiness at stake according to the different moral theories. Are there any other factors that should affect the initial distribution of control rights to theory representatives? Although no other suggestions strike me as *prima facie* plausible, I also have no strong arguments for the claim that no further factors should be taken into account here in determining the initial endowments of control rights.

In the remainder of this dissertation, I will assume that the decision

¹These two senses of 'stakes sensitivity' are also delineated and contrasted in Greaves and Cotton-Barratt 2024, §6.

maker's credence distribution over moral theories should be the *only* determinant of the initial distribution of control rights to theory representatives in any given choice situation. Once again, however, I make this assumption largely for the sake of simplicity, and unless otherwise noted my broad-strokes claims about IMM in the rest of this dissertation will not depend on how we resolve this theoretical choice point.

Recall that in the **Philanthropy** choice situation, control rights over the decision maker's total fortune are to be distributed between the two theory representatives R_1 and R_2 . R_1 represents the moral theory T_1 in which the decision maker has 60% credence, and R_2 represents the moral theory T_2 in which the decision maker has 40% credence. In light of this credence distribution, how should control rights over the decision maker's total fortune be divided up between R_1 and R_2 ?

At least in theory, one could imagine a wide range of possible ways in which the distribution of control rights could supervene upon the decision maker's credence distribution. For instance, each representative's share of these control rights could be proportional to the square root of the decision maker's credence in the corresponding moral theory; or else it could be proportional to the cube of this credence, or the logarithm. Nonetheless, the most natural and *prima facie* plausible view is just that each representative's share of these control rights should be *equal* to the decision maker's credence in the corresponding moral theory. This is the view that I will assume in the remainder of this dissertation, although yet again I make this assumption

largely for the sake of simplicity.

Under these assumptions, R_1 is to be initially endowed with control rights over 60% of the decision maker's fortune, and R_2 is to be endowed with control rights over the remaining 40%.

2.2 Proportional division

After this initial endowment, R_1 and R_2 have the option to trade or make contracts with each other. Are any mutually beneficial trades or contracts available to the two representatives under these circumstances?

Philanthropy is a highly stylised thought experiment in which I have specified that the decision maker may donate to only two possible charities (one providing deworming pills, and the other supporting local soup kitchens). T_1 directs the decision maker to donate as much as possible to deworming, whereas T_2 directs her to donate as much as possible to soup kitchens. Under these assumptions, R_1 and R_2 's preferences over how the decision maker divides her fortune between these two charities are diametrically opposed, in the sense that any change in the distribution of donations that is supported by one of these theory representatives will always be opposed by the other.

One might think that this result is enough to rule out the possibility of R_1 and R_2 negotiating any mutually beneficial trades or contracts in **Philanthropy**. However, this inference would be too quick, insofar as it ignores the

possibility of an *intertemporal* contract in which one theory representative benefits now, in return for agreeing to repay the other theory representative in one or more choice situations that the morally uncertain decision maker might confront at some later point in time, by granting the other theory representative greater control over those future choice situations than she would otherwise have been entitled to.

For example, suppose that our philanthropist has promised to help her neighbour move house tomorrow, but has just found out that keeping this promise would prevent her visiting her sick friend in hospital. And according to the moral theory T_1 , our philanthropist ought to keep her promise, even if this means that she cannot visit her sick friend. By contrast, according to T_2 , our philanthropist ought to visit her sick friend, even if this means that she cannot keep her promise. Furthermore, according to T_1 , how our decision maker distributes her charitable donations in **Philanthropy** is much more important than whether she keeps her promise to her neighbour. By contrast, according to T_2 , how our decision maker distributes her charitable donations in **Philanthropy** is much less important than whether she visits her sick friend. In other words, whereas the theory representative R_1 cares about the distribution of donations in **Philanthropy** much more than it cares about how the philanthropist will behave tomorrow, by contrast the theory representative R_2 cares about the distribution of donations in **Philanthropy** much less than it cares about how the philanthropist will behave tomorrow.

Under this set of assumptions, it will be in both R_1 and R_2 's interests to

agree to an intertemporal contract in which R_1 gains greater control over how the decision maker distributes her charitable donations in **Philanthropy**, in return for R_2 gaining greater control over how the decision maker trades off her duties to her sick friend against her promise to her neighbour. For instance, the two representatives might agree for R_1 to have control over 80 rather than 60% of the decision maker's fortune in **Philanthropy**, in return for R_2 having total control over whether the decision maker helps her neighbour or visits her sick friend.

I will discuss intertemporal contracts in much greater detail in chapter 4 below. Until then, however, I want to defer consideration of these extra complexities, by temporarily ruling out the potential for intertemporal gains from trade in scenarios like **Philanthropy**. To that end, let me amend my specification of **Philanthropy** by stipulating that our decision maker believes with certainty she will never encounter any choice situations other than **Philanthropy** wherein T_1 and T_2 issue incompatible directives. This rules out any intertemporal gains from trade, since neither R_1 neither R_2 cares about which of these theory representatives will decide how the philanthropist should behave in any future choice situations after **Philanthropy**.

With this assumption onboard, we are finally in a position to evaluate which distribution of donations is most appropriate in **Philanthropy** according to IMM. Recall that R_1 is to be initially endowed with control rights over 60% of the decision maker's fortune, and R_2 is to be endowed with control rights over the remaining 40%. And recall that since R_1 and R_2 have

diametrically opposed preferences over how the decision maker should divide her fortune between deworming and soup kitchens, there are no opportunities for intratemporal gains from trade within the **Philanthropy** choice situation. Moreover, since I have stipulated that our decision maker is certain she will never encounter any other choice situations in which R_1 and R_2 disagree, there are also no opportunities for intertemporal gains from trade either.

Under these conditions, R_1 and R_2 will not agree to any trades or contracts in our economic model for the **Philanthropy** choice situation. Instead, R_1 and R_2 will each use their endowment of control rights in the manner recommended by the moral theory that they represent. In other words: R_1 will instruct the philanthropist to donate 60% of her fortune to deworming, and R_2 will instruct the philanthropist to donate the remaining 40% of her fortune to soup kitchens. Hence, IMM implies that 60% to deworming and 40% to soup kitchens is the most appropriate distribution of donations in **Philanthropy**. The **Philanthropy** case illustrates that according to IMM, it is sometimes appropriate for a morally uncertain decision maker to split her resources between the different options favoured by each of the moral theories in which she has positive credence.

An alternative option that the philanthropist might have considered would have been for her to donate all of her fortune to deworming – the charity favoured by the moral theory T_1 in which she has 60% credence. However, this option strikes many of us as intuitively inappropriate in **Philanthropy**,²

²See chapter 1, n. 12 above.

at least under our assumption that the choiceworthiness ‘returns to scale’ from funding deworming and soup kitchens are constant according to both of the moral theories T_1 and T_2 .³ Under that assumption, donating all of the money to deworming would seem to cede too much influence to a moral theory T_1 in which the philanthropist is not particularly confident, and too little influence to a moral theory T_2 in which the philanthropist has 40% credence. Affording each theory its ‘due influence’ over the philanthropist’s donations would seem to require her to donate at least some of her fortune to soup kitchens, even if she donates the majority to deworming.

Fortunately, this kind of donation splitting is exactly what IMM recommends in **Philanthropy**. In this scenario, although the philanthropist is only somewhat more certain in T_1 than she is in T_2 , T_1 ’s recommendations to the philanthropist are diametrically opposed to T_2 ’s. Under these circumstances, it seems intuitively appropriate for the philanthropist to use her resources in a way that compromises between T_1 and T_2 ’s competing directives. Thus, in this sort of case at the very least, IMM provides an intuitively attractive criterion of appropriateness.

³If these returns to scale were instead assumed to be strictly increasing, then the most appropriate option in **Philanthropy** could instead be for our philanthropist to donate all of her fortune to deworming or to soup kitchens.

2.3 Alternatives

Hearteningly, IMM honours our intuitions about **Philanthropy** much more faithfully than its main rival, MEC. Of course, MEC's implications in **Philanthropy** depend on exactly how we spell out the details of T_1 and T_2 's views about the choiceworthinesses of all of the possible donation distributions. Still, we can get a sense of how MEC approaches scenarios like **Philanthropy** by adopting some simplified, 'toy-model' assumptions about choiceworthiness under T_1 and T_2 . I'll try to make these new toy-model assumptions somewhat more plausible than the rather stark assumptions about choiceworthiness values in **Philanthropy** that I temporarily used to illustrate a rather different point (as vividly as possible) in §2.1 above. (From now on, forget about that initial set of stark assumptions from §2.1.)

Specifically, let's suppose that, according to T_1 , for each dollar that the philanthropist can donate to charity, funding deworming is always five times as choiceworthy as funding soup kitchens. On the other hand, according to T_2 , for each dollar that the philanthropist can donate to charity, funding soup kitchens is always five times as choiceworthy as funding deworming. Finally, let's suppose *arguendo* that choiceworthiness units can be intertheoretically compared between T_1 and T_2 .⁴ In fact, just for simplicity of graphical illus-

⁴My own view is that intertheoretic choiceworthiness comparisons are impossible. However, recall from §1.2 above that MEC is only usable under the assumption that choiceworthiness units *can* be compared between all of the moral theories in which one has positive credence. Hence, in order to study what an advocate of MEC will say about a case like **Philanthropy**, one has to assume *arguendo* that intertheoretic choiceworthiness

tration, let's assume that choiceworthiness values according to both T_1 and T_2 can be measured on a single, shared scale of moral value. The choiceworthiness of spending a dollar on deworming according to T_1 is equal to the choiceworthiness of spending a dollar on soup kitchens according to T_2 . And, thus, the choiceworthiness of spending a dollar on soup kitchens according to T_1 will be equal to the choiceworthiness of spending a dollar on deworming according to T_2 .

To put these assumptions in algebraic terms, we are just stipulating that if our philanthropist spends $d\%$ of her money on deworming, and $(100 - d)\%$ on soup kitchens, then the total choiceworthiness of her donations will be $[5 \times d] + [1 \times (100 - d)] = 100 + 4d$ according to T_1 , and $[1 \times d] + [5 \times (100 - d)] = 500 - 4d$ according to T_2 . Intertheoretic expected choiceworthiness as a function of d will therefore be $[0.6 \times (100 + 4d)] + [0.4 \times (500 - 4d)] = 260 + 0.8d$, as illustrated in figure 2.2.

Hence, under these assumptions, expected choiceworthiness is clearly maximised when d is as large as possible – and so MEC implies that it is most appropriate for the philanthropist to donate all of her money to deworming and nothing at all to soup kitchens. Under these assumptions, then, MEC is unfaithful to the intuition – honoured by IMM – that it is appropriate for the philanthropist to split her donations between deworming and soup kitchens in **Philanthropy**. In fact, this result seems *especially* implausible given our extra assumption that the amount of choiceworthiness at stake in

comparisons can in fact be made in this scenario.

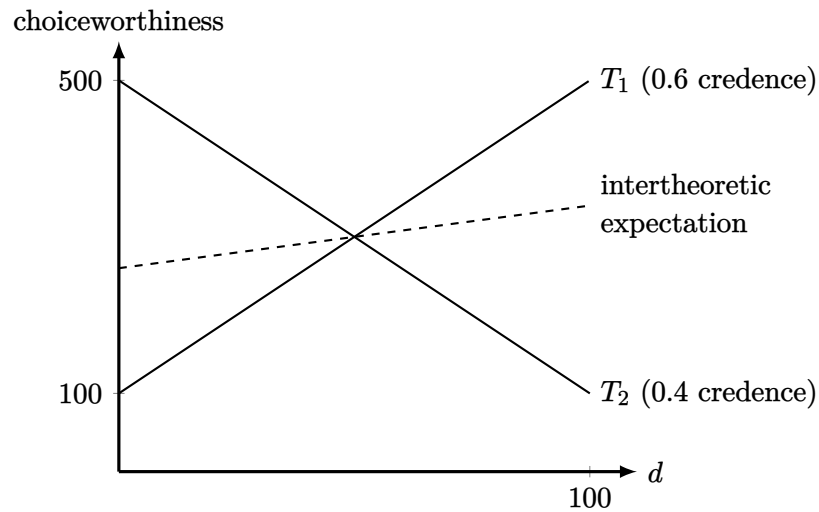


Figure 2.2: Choiceworthiness schedules in **Philanthropy**

Philanthropy according to T_1 is identical to the amount of choiceworthiness at stake according to T_2 . If we could say that the amount of choiceworthiness at stake in **Philanthropy** is much greater according to T_1 than it is according to T_2 , then we might find it somewhat more intuitively appropriate to donate all of the money to deworming, the charity favoured by T_1 . However, if T_1 and T_2 are in agreement about the amount of choiceworthiness at stake in **Philanthropy**, then no such defence is available for donating exclusively to deworming.

As I have already remarked, MEC's implications in **Philanthropy** depend on exactly how we spell out the details of T_1 and T_2 's views about the choiceworthinesses of all of the possible donation distributions.

In fact, given certain ways of spelling out these details, MEC can actually

sometimes imply that it is most appropriate for the philanthropist to split her fortune, donating some of it to deworming, and the rest to soup kitchens. For instance, suppose that for both T_1 and T_2 , marginal choiceworthiness returns are diminishing in the fraction of her fortune that the philanthropist donates to the moral theory's preferred charity. In other words, suppose that according to T_1 , the choiceworthiness difference between the philanthropist donating 0% versus 1% of her fortune to deworming is much greater than the choiceworthiness difference between the philanthropist donating 99% versus 100% of her fortune to deworming; and likewise *mutatis mutandis* for T_2 and soup kitchens. Under certain versions of this kind of diminishing marginal choiceworthiness in **Philanthropy**, MEC can imply that it is most appropriate for the philanthropist to split her fortune between deworming and soup kitchens.

Note that unlike the distribution recommended by IMM, there is no reason to expect that the donation distribution recommended by MEC will follow the same ratios as the philanthropist's credence distribution over moral theories. For instance, depending on how sharply marginal choiceworthiness diminishes according to T_1 and T_2 , perhaps MEC will imply that it is most appropriate for the philanthropist to split her fortune 80-20 between deworming and soup kitchens. Or, on the other hand, perhaps the appropriate distribution will be 50-50 instead. In short, although MEC can sometimes recommend splitting one's resources between the different options favoured by each of the moral theories in which one has credence, exactly how these re-

sources should be distributed will depend on the details of the moral theories in which one has credence.

How – if at all – are our intuitions about the appropriateness of resource splitting in **Philanthropy** sensitive to whether marginal choiceworthiness is diminishing according to T_1 and T_2 ? In my view, making the assumption that marginal choiceworthiness is diminishing for both T_1 and T_2 certainly strengthens the force of the intuition that it would be most appropriate for the philanthropist to split her donations between deworming and soup kitchens in **Philanthropy**. Not donating anything at all to soup kitchens would strike me as especially unreasonable and uncompromising under the assumption that according to T_1 , the choiceworthiness difference between donating 100% versus 99% to deworming is relatively small, whereas according to T_2 , the choiceworthiness difference between donating 0% versus 1% to soup kitchens is relatively large. Under these assumptions, the philanthropist donating not even 1 or 2% of her fortune to soup kitchens strikes me as an especially egregious example of depriving T_2 of its due influence over the philanthropist's donations.

Still, on the other hand, resource splitting also strikes me as intuitively appropriate in **Philanthropy** even under the assumption that marginal choiceworthiness is *non*-diminishing according to both T_1 and T_2 . Suppose we assume that according to both T_1 and T_2 , the choiceworthiness of the philanthropist's donation distribution in **Philanthropy** is linear in the fraction of her fortune d that she donates to deworming, as in figure 2.2 above. Even

under this assumption, MEC's recommendation to donate everything to deworming still strikes me as an inappropriately uncompromising 'winner takes all' response to **Philanthropy**. This response defers totally to a moral theory T_1 that the philanthropist still has significant doubts about, and not at all to a moral theory T_2 in which the philanthropist has quite a lot of credence.

The intuitive implausibility of MEC's resistance to compromise in cases of non-diminishing marginal choiceworthiness can be brought out particularly vividly by the following variation on **Philanthropy**:

Ninety-Nine Charities: some philanthropist is deciding where to donate her fortune. She faces a choice between ninety-nine different charities: C_1 through C_{99} . Suppose that this philanthropist has 1% credence in each of the ninety-eight different moral theories T_1 through T_{98} , and 2% credence in a final moral theory T_{99} . According to each moral theory T_n , the same-numbered charity C_n is the only charity that it would ever be choiceworthy for the philanthropist to donate to. In fact, each moral theory T_n implies that the choiceworthiness of any possible donation distribution is linearly proportional exactly and only to the amount that the philanthropist donates to the corresponding charity C_n . (Hence, marginal choiceworthiness is non-diminishing according to every moral theory T_1 through T_{99} .)

In IMM's bargaining model for **Ninety-Nine Charities**, T_1 through T_{99} will be represented by ninety-nine theory representatives R_1 through R_{99} ,

who will bargain with each other to determine how the philanthropist should behave. R_1 through R_{98} will each be initially endowed with control rights over 1% of the philanthropist's fortune, whereas R_{99} will have 2% (because the philanthropist has 2% credence in the corresponding moral theory T_{99}). As we did for **Philanthropy**, let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for intertemporal gains from trade or contract. Moreover, let us also assume that the choiceworthiness returns to scale from funding any one of the ninety-nine charities C_1 through C_{99} are *constant* according to all of the moral theories T_1 through T_{99} .

Under these assumptions, IMM implies that it is most appropriate for the philanthropist in **Ninety-Nine Charities** to donate 1% of her fortune to each of the charities C_1 through C_{98} , and the remaining 2% of her fortune to the final charity C_{99} . As in **Philanthropy**, none of the theory representatives R_1 through R_{99} have anything to gain from trading or making contracts with any of the other theory representatives. Hence, each representative R_n will simply instruct the philanthropist to donate its share of her fortune to R_n 's favoured charity C_n .

This strikes me as an intuitively attractive response to **Ninety-Nine Charities**. In this choice situation, the philanthropist is certain that there is only one charity that it would ever be choiceworthy for her to donate to, but is radically uncertain about which of the ninety-nine charities C_1 through

C_{99} is the choiceworthy one. Donating some of her fortune to each of the charities C_1 through C_{99} leaves the philanthropist certain that at least some of her fortune will have been donated to a worthwhile charity. In light of the philanthropist's radical moral uncertainty, this strikes me as the best available compromise between the competing directives of T_1 through T_{99} .

Once again, these intuitions are honoured by IMM, but are inconsistent with MEC. Suppose *arguendo* that choiceworthiness units can be intertheoretically compared between all of the moral theories T_1 through T_{99} , and that the amount of choiceworthiness at stake in **Ninety-Nine Charities** according to any of these moral theories is roughly equal to the amount at stake according to any of the other theories.

Under these assumptions, MEC will have the counterintuitive implication that it is most appropriate for the philanthropist to donate her entire fortune to C_{99} in **Ninety-Nine Charities**. For each dollar in the philanthropist's fortune, the expected choiceworthiness of donating that dollar to C_{99} will be – *ex hypothesi* – roughly twice as great as the choiceworthiness of donating that dollar to any of the other charities C_1 through C_{98} . Hence, expected choiceworthiness will be maximised by donating every dollar in the philanthropist's fortune to C_{99} . And so MEC implies that this is the most appropriate option in **Ninety-Nine Charities**.

However, this donation distribution recommended by MEC strikes many of us as a highly inappropriate choice in **Ninety-Nine Charities**. Donating all of her money to C_{99} would leave the philanthropist 98% certain that all of

her donations were completely unchoiceworthy. This course of action would seem to cede too much influence to a moral theory T_{99} in which the philanthropy only has 2% credence, and too little influence to the moral theories T_1 through T_{98} , which together account for fully 98% of the philanthropist's credence over moral theories. Affording each moral theory its 'due influence' over the philanthropist's donations would seem to require somehow splitting her fortune between the ninety-nine charities C_1 through C_{99} that she thinks might be choiceworthy to donate to – something which MEC does not recommend.

Overall, then, IMM seems to honour our intuitions in favour of compromise in resource-splitting choice situations much more faithfully than its main rival MEC.

In fact, IMM also has this advantage over several of its other rivals as well, including the '*My Favourite Theory*' (henceforth: 'MFT') and '*My Favourite Option*' (henceforth: 'MFO') criteria of appropriateness – both of which have been discussed in the literature on moral uncertainty as potential competitors to MEC. According to MFT:

(MFT) some option A is appropriate in some choice situation S
iff according to the moral theory in which the decision maker has
greatest credence, A is the most choiceworthy option available in
S.⁵

And according to MFO:

⁵Gracely 1996; Gustafsson and Torpman 2014.

(MFO) some option A is appropriate in some choice situation S iff under the decision maker's credence distribution over moral theories, A is the option that has the highest probability of being the most choiceworthy option available in S .⁶

The decision maker in **Philanthropy** has 60% credence in the moral theory T_1 , according to which the philanthropist donating her entire fortune to deworming is the most choiceworthy option available in S . Hence, MFT and MFO both imply that this is also the most appropriate option in **Philanthropy**.

Similarly, the decision maker in **Ninety-Nine Charities** has 2% credence in the moral theory T_{99} , according to which C_{99} is the most choiceworthy charity; whereas she has at most 1% credence in any moral theory other than T_{99} , and hence at most 1% credence in any alternative to C_{99} being the most choiceworthy charity. Thus, MFT and MFO both imply that the philanthropist donating her entire fortune to C_{99} is the most appropriate option in **Ninety-Nine Charities**.

Thus, in these two scenarios (**Philanthropy** and **Ninety-Nine Charities**), MFT and MFO both agree with MEC that it would be inappropriate for the philanthropist to split her donations over several different charities. This uncompromising, 'winner takes all' response to these scenarios strikes me as an important disadvantage of MEC, MFT and MFO, as compared

⁶MFO has been discussed (without endorsement) by Lockhart 2000, chapters 2-4; Gustafsson and Torpman 2014, §2; MacAskill and Ord 2020, §4.

against the IMM approach.

2.4 Contracts

I have now spent quite a lot of time discussing the **Philanthropy** scenario, and its close cousin **Ninety-Nine Charities**. But **Philanthropy** is just one example of a choice situation in which the decision maker has to decide how to use some endowment of a finite and continuously divisible resource like money or time. In **Philanthropy**, gains from trade or contract are unavailable to the theory representatives R_1 and R_2 (recall §2.2 above). However, there are many other resource distribution choice situations in which gains from trade or contract are available.

We can illustrate the possibility of gains from contract using the following variation on **Philanthropy**:

Three Charities: some philanthropist is deciding where to donate her fortune. She faces a choice between three charities: A, B, and C. Suppose that this philanthropist has 50% credence in the moral theory T_1 , and 50% credence in T_2 . According to T_1 , it is highly choiceworthy to donate to A, almost as choiceworthy to donate to B, but scarcely choiceworthy at all to donate to C. Conversely, according to T_2 , it is highly choiceworthy to donate to C, almost as choiceworthy to donate to B, but scarcely choiceworthy at all to donate to A (illustrated in figure 2.3).

	T ₁ : 50% credence	T ₂ : 50% credence
↑ increasing choiceworthiness	A	C
	B	B
	C	A

Figure 2.3: Choiceworthiness schedules in **Three Charities**

In IMM's bargaining model for **Three Charities**, T_1 and T_2 will once again be represented by two theory representatives, R_1 and R_2 , who this time around will each be initially endowed with control rights over 50% of the philanthropist's fortune. As we did for **Philanthropy** and **Ninety-Nine Charities**, let us again assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for intertemporal gains from trade or contract. Moreover, let us also assume that the choiceworthiness returns to scale from funding any one of the charities A, B, or C are *constant* according to both of the moral theories T_1 and T_2 .

Recall that in **Philanthropy** and **Ninety-Nine Charities**, the theory representatives R_1 and R_2 have diametrically opposed preferences over how the decision maker should divide her fortune between the available charities, in the sense that any change in the distribution of donations that is supported by one of these theory representatives will always be opposed by the other.

In **Three Charities**, however, R_1 and R_2 do not have diametrically opposed preferences. Instead, the introduction of a third charity (B) that both theory representatives regard as somewhat choiceworthy opens upon the possibility for R_1 and R_2 to agree to a mutually beneficial contract in **Three Charities**.

Imagine, first of all, an outcome in which R_1 and R_2 do not make any contracts with each other. Under these conditions, it will be in R_1 's best interests to instruct the philanthropist to donate R_1 's half of her fortune to charity A; and it will be in R_2 's best interests to instruct the philanthropist to donate R_2 's half of her fortune to charity C. Unfortunately, neither R_1 nor R_2 would be particularly happy with this outcome, since each would think that the philanthropist is wasting half of her fortune on the charity recommended by the other theory representative.

Now imagine, however, an outcome in which R_1 and R_2 both agree to instruct the philanthropist to donate to charity B. Since R_1 and R_2 both think that donating to B is almost as choiceworthy as donating to their favoured charity, it is safe to assume that both theory representatives will regard this outcome as a significant improvement over the outcome in which the philanthropist splits her donations 50-50 between charities A and C. In other words, it is safe to assume that a 50-50 split between A and C is *strictly Pareto dominated* by (among other things) the philanthropist donating all of her fortune to charity B. Since no rational economic agents with the ability to make contracts would ever settle for a strictly Pareto-dominated outcome, IMM thus implies that the philanthropist splitting her donations 50-50 be-

tween charities A and C would be inappropriate in **Three Charities**.

This result raises the question of which donation distribution is *most* appropriate in **Three Charities**. Would it be most appropriate for the philanthropist to donate all of her fortune to charity B? Or, would it be most appropriate for her to donate most of her fortune to B, but some to A and/or C?

At the moment, we are missing several pieces of information that we would need in order to answer these questions definitively. Firstly, I have not yet specified in any detail how IMM should model the process of bargaining between the theory representatives. If there are multiple possible contracts that would all be Pareto improvements over a 50-50 split between A and C, then which of these contracts will R_1 and R_2 settle on after bargaining with each other? I will discuss this question in §7 below.

There is also some other information that we do not yet have, but would almost certainly need if we wanted to work out which donation distribution is most appropriate in **Three Charities** – *viz.* information about exactly how T_1 and T_2 compare the choiceworthinesses of donating to A, B, and C. At the moment, my specification of **Three Charities** is compatible with, for instance, R_1 being somewhat happier than R_2 to compromise on charity B. For example, T_1 might imply that donating an extra dollar to B is always 90% as choiceworthy as donating an extra dollar to A, whereas T_2 might imply that donating an extra dollar to B is only 70% as choiceworthy as donating an extra dollar to C. *Prima facie*, it strikes me as quite plausible

to suppose that IMM should be sensitive to these kinds of details. If R_2 would be less happy than R_1 – in the sense that I have just stipulated – with a contract to replace the ‘default’ 50-50 distribution over A and C with an instruction for the philanthropist to donate exclusively to B, then perhaps the most appropriate distribution would be for the philanthropist to donate *most* of her money to B, but also a small amount to C, as a way to sweeten the deal for R_2 .

Whether IMM endorses this suggestion will depend on exactly how IMM models the details of the bargaining between theory representatives. As I have already mentioned, I will discuss this theoretical choice point in §7 below. However, I will not try to conclusively settle that choice point in this dissertation. In other words, I won’t try to prove that the bargaining model I suggest in §7 is necessarily the best bargaining model to use with IMM. Rather, my more modest aim will just be to show that there is at least one bargaining model under which IMM is more attractive than rival theories of appropriateness like MEC – and hence to show that IMM represents a significant theoretical step in the right direction.

Returning to the question of which donation distribution is most appropriate in **Three Charities**, I have suggested that we could answer this question definitively only if we knew (i) how IMM should model inter-representative bargaining, and (ii) exactly how T_1 and T_2 compare donations to charities A, B, and C. Since I will defer discussion of inter-representative bargaining until §7 of this dissertation, my definitive response to **Three Charities** will

also have to wait until then.

This point notwithstanding, there are nonetheless some ways of stipulating exactly how T_1 and T_2 compare donations to charities A, B, and C under which *any* plausible model of inter-representative IMM bargaining will imply that the most appropriate option in **Three Charities** is for the philanthropist to donate everything to charity B. In particular, I have in mind cases in which R_1 's bargaining position vis-à-vis R_2 is in a certain sense '*structurally identical*' to R_2 's bargaining position vis-à-vis R_1 .

For example, suppose that according to T_1 , donating an extra dollar to B is always 90% as choiceworthy as donating an extra dollar to A, and donating an extra dollar to C is only 10% as choiceworthy. Symmetrically, suppose that according to T_2 , donating an extra dollar to B is always 90% as choiceworthy as donating an extra dollar to C, and donating an extra dollar to A is only 10% as choiceworthy.

Under these assumptions, R_1 and R_2 's bargaining positions are structurally identical, in the sense that there is some description of R_1 's bargaining position vis-à-vis R_2 that (i) captures every fact that could plausibly be relevant to the IMM bargaining between R_1 and R_2 , but that (ii) is also a similarly complete description of R_2 's bargaining position vis-à-vis R_1 . In particular, R_1 and R_2 's bargaining positions can both be described as follows:

1. The representative is initially endowed with control over 50% of the philanthropist's fortune, and the representative's 'adversary' is initially endowed with control over the other 50%.

2. There are three charities that the philanthropist can donate to. The representative thinks that one of these charities is the best, another is 90% as good as the best, and the final charity is only 10% as good as the best.

3. Finally, the adversary's preference ordering over these three charities reverses the representative's. Exactly like the representative, however, the adversary also thinks that the middle-ranked charity is 90% as good as the best, and the worst is only 10% as good as the best.

Thus, in every respect that could plausibly make a difference to IMM's bargaining process, R_1 's bargaining position is identical to R_2 's.

Under these conditions, no plausible specification of IMM's bargaining process could privilege either of the two theory representatives in **Three Charities** over the other representative. For instance, no plausible version of IMM could privilege R_1 over R_2 by implying that it is most appropriate for the philanthropist to donate most of her fortune to charity B, and all of the rest to charity A. Since R_1 and R_2 's bargaining positions are structurally identical to each other, there could be no reasonable basis for privileging R_1 's objectives over R_2 's like this.

Thus, under the assumption that R_1 and R_2 's bargaining positions are structurally identical, any plausible version of IMM will imply that the most appropriate donation distribution in **Three Charities** assigns no more and no less to charity A than it assigns to charity C.

Furthermore, our new assumptions about T_1 and T_2 also imply that switching any \$1 of the philanthropist's fortune from A to B at the same time as switching another \$1 of her fortune from C to B will always be a strong Pareto improvement, in the sense that it will make both R_1 and R_2 happier than they otherwise would have been. Thus, any response to **Three Charities** that assigns the same nonzero amount of money to both A and C must be Pareto dominated by the philanthropist donating all of her fortune to B. And so given that no rational bargainers would ever settle for a Pareto-dominated outcome, and given my new assumptions about T_1 and T_2 , any plausible version of IMM will imply that a response to **Three Charities** that assigns the same *nonzero* amount of money to both A and C must be inappropriate.

To summarize: under the assumption of structurally identical bargaining positions, the most appropriate donation distribution in **Three Charities** must assign an amount to A equal to the amount that it assigns to C; and, furthermore, both of these amounts must be *zero*. Hence, assuming structurally identical bargaining positions, any plausible version of IMM will imply that it is most appropriate for the philanthropist to donate her entire fortune to charity B in **Three Charities**.

This discussion of **Three Charities** illustrates that IMM's prescriptions in scenarios where there are potential gains from contract are importantly different from IMM's prescriptions in cases like **Philanthropy** where there are no potential gains from trade or contract. Recall that in **Philanthropy**,

IMM implies that it is most appropriate for the philanthropist to split her donations between the different options favoured by each of the moral theories in which the philanthropist has positive credence. By contrast, in versions of **Three Charities** where we assume structurally identical bargaining positions, any plausible version of IMM will imply that it is most appropriate for the philanthropist to donate her entire fortune to a compromise charity, that is most favoured by neither of the moral theories in which the philanthropic has positive credence.

Underwriting this kind of ‘hedging’ in scenarios like **Three Charities** is an attractive implication that IMM has in common with its chief rival, MEC. Suppose *arguendo* that choiceworthiness units can be intertheoretically compared between T_1 and T_2 , and that the amount of choiceworthiness at stake in **Three Charities** according to T_1 is at least roughly equal to the amount at stake according to T_2 . Under these assumptions, MEC will agree with IMM that it is most appropriate for the philanthropist to donate her entire fortune to charity B in **Three Charities**.

Of course, MEC’s rationale for hedging in **Three Charities** is importantly different from IMM’s. According to MEC, donating everything to B is the most appropriate option in **Three Charities** because this option is guaranteed to be somewhat choiceworthy, and so – unlike donating to charities A or C – does not carry any risk of being highly unchoiceworthy. By contrast, according to IMM, donating everything to B is the most appropriate option because it is the best and most balanced compromise between T_1

and T_2 's competing directives to donate to charities A and C respectively.

Some of MEC's leading advocates have suggested that underwriting hedging in scenarios like **Three Charities** is an important advantage of MEC over some of its rivals like MFT and MFO.⁷ In **Three Charities**, none of the moral theories in which the philanthropist has positive credence imply that donating to charity B is maximally choiceworthy. Hence, MFT and MFO both imply that donating to charity B cannot be the most appropriate option in **Three Charities**. That IMM also avoids this unattractive implication is thus an important advantage of IMM over alternative competitors to MEC like MFT and MFO, and it undercuts an important part of the MEC advocates' purported case in favour of MEC as the best of all possible approaches to moral uncertainty.

2.5 Trade

Recall that in IMM's models of the **Philanthropy** and **Ninety-Nine Charities** scenarios, there were no gains from trade or contract available to any of the theory representatives, and so none of the theory representatives would agree to any trades or contracts with each other in post-endowment bargaining phases of IMM's models of these choice situations. In **Three Charities**, by contrast, the introduction of a third, 'compromise' charity B created new possibilities for inter-representative gains from contract, and so we know that

⁷Lockhart 2000, chapter 4; MacAskill, Bykvist and Ord 2020, p. 45; MacAskill and Ord 2020, pp. 337-8.

the theory representatives R_1 and R_2 will agree to some kind of contract with each other in the bargaining phase of IMM's model for this choice situation.

However, introducing a new compromise charity is not the only way to modify **Philanthropy** that will have the effect of inducing R_1 and R_2 to strike a bargain with each other. Another way to stimulate bargaining is to create new possibilities for inter-representative gains from trade by introducing a second *resource*, in addition to the philanthropist's monetary fortune. For instance, consider the following choice situation:

Double Distribution: some decision maker is choosing (i) where to donate her fortune, and (ii) what to do with her free time. This decision maker must distribute her fortune between two charities: the first of which provides deworming pills to distant children, and the second of which supports local soup kitchens. Similarly, our decision maker must also distribute her free time between two possible uses: the first of which is campaigning for nuclear disarmament, and the second of which is volunteering at a local orphanage. Suppose that this decision maker has, say, 50% credence in the moral theory T_1 , according to which she should donate as much of her money as possible to deworming, and as much of her time as possible to nuclear disarmament. She also has 50% credence in the moral theory T_2 , according to which she should donate as much of her money as possible to soup kitchens, and as much of her time as possible to the local orphanage.

In IMM's bargaining model for **Double Distribution**, T_1 and T_2 will once again be represented by two theory representatives, R_1 and R_2 , who will each be initially endowed with control rights over 50% of the decision maker's monetary fortune, and also over 50% of her free time. As we always do, let us again assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contracts involving other scenarios in addition to **Double Distribution**. Moreover, let us also assume that the returns to scale from each of the possible uses for time and money available in **Double Distribution** are *constant*, according to both of the moral theories T_1 and T_2 . Finally, for the sake of simplicity, for the moment I will assume that the decision maker knows exactly how much money she has in her fortune and exactly how much free time she has available. (I will discuss some of the complications introduced by relaxing this assumption in chapters 3 and 4 below.)

Recall that in **Philanthropy** and **Ninety-Nine Charities**, the theory representatives R_1 and R_2 have diametrically opposed preferences over how the decision maker should divide her fortune between the available charities: any change in the distribution of donations that is supported by either one of these theory representatives would always be opposed by the other. This diametric opposition between R_1 and R_2 's preferences over the decision maker's charitable donations also obtains in **Double Distribution** – provided we *hold all else equal*. In other words, *holding constant* how the decision maker

uses her free time, any change in the distribution of donations supported of either one of these theory representatives will always be opposed by the other. Symmetrically, in **Double Distribution** R_1 and R_2 also have diametrically opposed preferences over how the decision maker should divide her time between its two available uses. Holding constant how the decision maker uses her monetary fortune, any change in the distribution of volunteering that is supported by one of these theory representatives will always be opposed by the other.

Thus, in **Double Distribution**, R_1 and R_2 have diametrically opposed preferences over how each of the decision maker's two different resource endowments (time and money) should be distributed, holding constant how the decision maker uses her endowment of the other of these resources. However, this result is totally compatible with there being gains from trade available to R_1 and R_2 , because it is compatible with R_1 and R_2 having preferences that are not diametrically opposed to each other over how the decision maker's two resource endowments should be *jointly* distributed. More particularly, gains from trade will be available to R_1 and R_2 in **Double Distribution** iff R_1 and R_2 disagree with each other about, roughly speaking, how the choiceworthiness value of money compares to the choiceworthiness value of time (or, more precisely: how the choiceworthiness value of control over one unit of the decision maker's monetary fortune compares against the choiceworthiness value of control over one unit of the decision maker's time).

For instance, suppose that according to T_1 , the total choiceworthiness of

resource:	money	time
R_1 's preferred use:	deworming	nuclear disarmament
R_2 's preferred use:	soup kitchens	orphanage
representative that cares about this resource more than the other resource:	R_2	R_1

Figure 2.4: R_1 and R_2 's preferences in **Double Distribution**

the decision maker's overall response to **Double Distribution** is much less sensitive to how she distributes her charitable donations than it is to how she uses her free time. By contrast, according to T_2 , total choiceworthiness in **Double Distribution** is much more sensitive to charitable donations than it is to how the decision maker uses her free time. In other words, R_1 cares about money less than she cares about time, whereas R_2 cares about money more than she cares about time. Figure 2.4 summarizes what we know about R_1 and R_2 under these assumptions.

Imagine, first of all, an outcome in which R_1 and R_2 do not make any trades or contracts with each other. Under these conditions, it will be in R_1 's best interests to instruct the decision maker (i) to donate R_1 's half of her monetary fortune to deworming, and (ii) to use R_1 's half of her free time campaigning for nuclear disarmament. At the same time, it will be in R_2 's best interests to instruct the decision maker (i) to donate R_2 's half of her fortune to soup kitchens, and (ii) to use R_2 's half of her free time working at the orphanage. Unfortunately, neither R_1 nor R_2 would be particularly

happy with this outcome. R_1 would be unhappy mostly because R_1 would think that the decision maker is wasting half of her free time working at the orphanage, whereas R_2 would be unhappy mostly because R_2 would think that the decision maker is wasting half of her monetary fortune on deworming.

Now imagine, however, an outcome in which R_1 and R_2 both agree to trade parts of their initial resource endowments. In particular, imagine that R_1 transfers to R_2 part or all of R_1 's initial endowment of control rights over the decision maker's monetary fortune, in return for R_2 transferring to R_1 part or all of R_2 's initial endowment of control rights over the decision maker's free time. After this trade, R_1 will then instruct the decision maker (i) to donate to deworming any part of the decision maker's monetary fortune that R_1 still has control over post-trade, and (ii) to campaign for nuclear disarmament using all of R_1 's post-trade share of the decision maker's free time. At the same time, R_2 will instruct the decision maker (i) to donate to soup kitchens all of R_2 's post-trade share of the decision maker's monetary fortune, and (ii) to work at the orphanage using any part of the decision maker's free time that R_2 still has control over post-trade.

Remember, R_1 thinks that how the decision maker uses her free time is a 'higher stakes' choice than how the decision maker allocates her charitable donations, whereas R_2 thinks the opposite. Hence, it is safe to assume that at least some trades of this form will be Pareto improvements over the 'no trade' outcome in which the decision maker splits both of her resource endowments 50-50 between R_1 and R_2 's preferred uses for them. In other

words, it is safe to assume that the ‘no trade’ outcome is Pareto dominated by at least some outcomes in which R_1 has greater than 50% control over the decision maker’s time, and R_2 has greater than 50% control over the decision maker’s money. Since no rational economic agents with the ability to make contracts would ever settle for a Pareto-dominated outcome, IMM thus implies that the decision maker splitting both of her resource endowments 50-50 between R_1 and R_2 ’s preferred uses for them would be inappropriate in **Double Distribution**.

Which joint distribution of donations and use of free time would be *most* appropriate in **Double Distribution**? All I have really said so far is that soup kitchens should get more donations than deworming, and disarmament campaigning should get more time than the orphanage. So for all I have said so far, it could be most appropriate for the decision maker to donate *all* of her money to soup kitchens, and spend *all* of her free time on disarmament campaigning. Or – rather less cleanly – perhaps it could be most appropriate for the decision maker to (i) donate all of her monetary fortune to soup kitchens, and (ii) split her free time 80-20 between campaigning for nuclear disarmament and volunteering at the orphanage. Or, the flipside, perhaps it could instead be most appropriate for the decision maker to (i) split her monetary fortune 30-70 between deworming and soup kitchens, and (ii) use all of her free time campaigning for nuclear disarmament. Finally, perhaps it could in fact be most appropriate for the decision maker to divide both her money and her time between their two possible uses, donating most but

not all of her money to soup kitchens, and using most but not all of her time campaigning for nuclear disarmament.

At the moment, we are once again (recall §2.4 above) missing several pieces of information that we would need in order to answer definitively which overall response to **Double Distribution** is most appropriate. Firstly, I have not specified in any detail how IMM should model the process of bargaining between the theory representatives. If there are multiple possible trades that would all be Pareto improvements over the ‘no trade’ outcome, then which of these trades will R_1 and R_2 settle on after bargaining with each other? I will discuss this question in §7 below.

There is also some other information that we do not yet have, but would almost certainly need if we wanted to work out which overall response to **Double Distribution** is most appropriate – *viz.* information about exactly how T_1 and T_2 each compare the choiceworthinesses of donating to deworming, donating to soup kitchens, disarmament campaigning, and orphanage volunteer work. At the moment, my specification of **Double Distribution** is compatible with, for instance, R_1 being somewhat more eager than R_2 to trade R_1 ’s control over money for R_2 ’s control over free time. For example, T_1 might imply that switching an hour of free time from orphanage volunteering to disarmament campaigning always increases choiceworthiness by 10 times as much as switching \$100 of donations from soup kitchens to deworming, whereas T_2 might imply that switching \$100 of donations from deworming to soup kitchens is only 5 or 6 times more choiceworthy than switching an hour

of free time from disarmament campaigning to orphanage volunteering.

Prima facie, it strikes me as quite plausible to suppose that IMM should be sensitive to these kinds of details. If R_2 would be less delighted than R_1 – in the sense that I have just stipulated – with a possible outcome in which the decision maker (i) donates all of her money to soup kitchens but (ii) uses all of her free time to campaign for disarmament, then perhaps the most appropriate distribution would require the decision maker (i) to donate all of her money to soup kitchens, but would also require her (ii) to spend a small fraction of her free time volunteering at the orphanage instead of campaigning for disarmament, as a way to sweeten the deal for R_2 .

Whether IMM endorses this suggestion will depend on exactly how IMM models the details of the bargaining between theory representatives. As I have already mentioned, I will discuss this theoretical choice point in §7 below.

Returning to the question of which overall response to **Double Distribution** is most appropriate, I have suggested that we could answer this question only if we knew (i) how IMM should model inter-representative bargaining, and (ii) the details of T_1 and T_2 's views about choiceworthiness in **Double Distribution**. Since I will defer discussion of inter-representative bargaining until §7 of this dissertation, my definitive response to **Double Distribution** will also have to wait until then.

This point notwithstanding, there are nonetheless some ways of stipulating T_1 and T_2 's views about choiceworthiness in **Double Distribution**

under which *any* plausible model of inter-representative IMM bargaining will imply that the most appropriate option in **Double Distribution** is for the decision maker to donate all of her monetary fortune to soup kitchens, and use all of her free time campaigning for nuclear disarmament. Once again (recall §2.4 above), I have in mind cases in which R_1 's bargaining position vis-à-vis R_2 is structurally identical to R_2 's bargaining position vis-à-vis R_1 .

For example, suppose that according to T_1 , switching some proportion $x\%$ of the decision maker's total free time from orphanage volunteering to disarmament campaigning always increases choiceworthiness by 10 times as much as switching the same proportion ($x\%$) of her monetary fortune from soup kitchens to deworming. Symmetrically, suppose that according to T_2 , switching some proportion $x\%$ of the decision maker's monetary fortune from deworming to soup kitchens always increases choiceworthiness by 10 times as much as switching the same proportion ($x\%$) of her total free time from disarmament campaigning to orphanage volunteering.

Under these conditions, R_1 and R_2 's bargaining positions are structurally identical to each other in the sense defined in §2.4 above, because R_1 and R_2 's bargaining positions can both be comprehensively described as follows:

1. The representative is initially endowed with control over 50% of the decision maker's monetary fortune, and 50% of her free time – as is the representative's 'adversary.'
2. There are two possible uses for each of the two resources, money and

time. And the representative thinks that for each of these two resources, one of the two possible uses is better than the other.

3. Unfortunately, however, the adversary disagrees with the representative about which of the two uses is best for both of these two resources.
4. The representative also cares about one of these two resources more than she cares about the other. Moving any $x\%$ of the decision maker's total stock of that resource from its worst to its best use always makes ten times as much of a difference to the representative as moving $x\%$ of the decision maker's total stock of the other resource from its worst to its best use.
5. Finally, the adversary disagrees with the representative about which of these two resources is most important. Exactly as for the representative, however, moving any $x\%$ of the decision maker's total stock of the resource that the adversary cares the most about from its worst to its best use makes ten times as much of a difference to the adversary as moving $x\%$ of the decision maker's total stock of the other resource from its worst to its best use.

Hence, in every respect that could plausibly make a difference to IMM's bargaining process, R_1 's bargaining position vis-à-vis R_2 is identical to R_2 's bargaining position vis-à-vis R_1 .

Under these conditions, no plausible specification of IMM's bargaining process could privilege either of these two theory representatives in **Dou-**

ble Distribution over the other representative. For instance, no plausible version of IMM could privilege R_1 over R_2 by implying that it is most appropriate for the decision maker to spend all of her time campaigning for nuclear disarmament, but also donate at least some of her money to deworming as opposed to soup kitchens. Since R_1 and R_2 's bargaining positions are structurally identical to each other, there could be no reasonable basis for privileging R_1 's objectives over R_2 's like this.

Thus, under the assumption that R_1 and R_2 's bargaining positions are structurally identical, any plausible version of IMM will imply that the most appropriate response to **Double Distribution** assigns a proportion of the decision maker's time to R_1 's preferred use (disarmament campaigning) that is equal to the proportion of the decision maker's money that it assigns to R_2 's preferred use (soup kitchens).

Furthermore, our new assumptions about T_1 and T_2 also imply that switching any $x\%$ of the decision maker's money from deworming to soup kitchens at the same time as switching any $x\%$ of her time from orphanage volunteering to disarmament campaigning will always be a strong Pareto improvement, in the sense that it will make both R_1 and R_2 happier than they otherwise would have been. Thus, any response to **Double Distribution** that assigns a nonzero proportion of the decision maker's money to deworming and the same nonzero proportion of her time to orphanage volunteering must be Pareto dominated by the decision maker spending all of her time on disarmament campaigning and all of her money on soup kitchens. And so

given that no rational bargainers would ever settle for a Pareto-dominated outcome, and given my new assumptions about T_1 and T_2 , any plausible version of IMM will imply that a response to **Double Distribution** that assigns a nonzero proportion of the decision maker's money to deworming and the same *nonzero* proportion of her time to orphanage volunteering must be inappropriate.

To summarize: under the assumption of structurally identical bargaining positions, the most appropriate response to **Double Distribution** must assign a proportion of time to disarmament campaigning equal to the proportion of money that it assigns to soup kitchens; and, furthermore, both of these proportions must be 100%. Hence, assuming structurally identical bargaining positions, any plausible version of IMM will imply that it is most appropriate for the decision maker to spend all of her time on disarmament campaigning, and donate all of her money to soup kitchens.

This discussion of **Double Distribution** illustrates that IMM's prescriptions in scenarios where there are potential gains from trade are importantly different from IMM's prescriptions in cases like **Philanthropy** where there are no potential gains from trade or contract. Recall that in **Philanthropy**, IMM implies that it is most appropriate for R_1 's initial share of the philanthropist's resources to be used for R_1 's favoured purpose, and likewise *mutatis mutandis* for R_2 . By contrast, in versions of **Double Distribution** where we assume structurally identical bargaining positions, any plausible version of IMM will imply that it is most appropriate for the decision maker

to donate all of her money to R_2 but not R_1 's favoured use for it (soup kitchens), and for her to spend all of her time on R_1 but not R_2 's favoured use for it (disarmament campaigning).

This response to **Double Distribution** strikes me as *prima facie* intuitively plausible. In deciding how to distribute money, IMM recommends that the decision maker should be guided most strongly by the moral theory T_2 according to which what is at stake in distributing money is greater than what is at stake in allocating free time. Similarly, in deciding how to allocate free time, IMM recommends that the decision maker should be guided most strongly by the moral theory T_1 according to which what is at stake in allocating free time is greater than what is at stake in distributing money. This strikes me as an attractive response to **Double Distribution**.

Underwriting this kind of 'stake sensitivity' in scenarios like **Double Distribution** is an attractive implication that IMM has in common with its chief rival, MEC. Suppose *arguendo* that choiceworthiness units can be intertheoretically compared between T_1 and T_2 , and that the total amount of choiceworthiness at stake in **Double Distribution** according to T_1 is at least roughly equal to the amount at stake according to T_2 . Under these assumptions, MEC will agree with IMM that it is most appropriate for the decision maker to spend all of her time on disarmament campaigning, and donate all of her money to soup kitchens.

Of course, MEC's rationale for stake sensitivity in **Double Distribution** is importantly different from IMM's. According to MEC, disarmament

campaigning plus donating to soup kitchens is the most appropriate option in **Double Distribution** because disarmament campaigning represents a more favourable gamble than orphanage volunteering, and soup kitchens represent a more favourable gamble than deworming. By contrast, according to IMM, disarmament campaigning plus donating to soup kitchens is the most appropriate option in **Double Distribution** because it is the best and most balanced compromise between T_1 and T_2 's conflicting directives in this choice situation.

Some of MEC's leading advocates have suggested that underwriting some kind of stakes sensitivity in scenarios like **Double Distribution** is an important advantage of MEC over some of its rivals like MFT and MFO.⁸ In **Double Distribution**, none of the moral theories in which the decision maker has positive credence imply that it is maximally choiceworthy for her to spend all of her time on disarmament campaigning and donate all of her money to soup kitchens. Hence, MFT and MFO both imply that this cannot be the most appropriate response to **Double Distribution**. That IMM avoids this unattractive implication is thus an important advantage of IMM over alternative competitors to MEC like MFT and MFO, and it undercuts an important part of the MEC advocates' purported case in favour of MEC as the best of all possible approaches to moral uncertainty.

⁸Sepielli 2009, p. 11; 2010; MacAskill, Bykvist and Ord 2020, pp. 44-5; MacAskill and Ord 2020, p. 337.

Chapter 3

Risk

In the real world, almost all morally uncertain decision makers are also often *descriptively* uncertain about exactly what the consequences would be of the various possible choices open to them. For example, doctors might be uncertain about whether a drug will harm or benefit their patients; and policymakers might be uncertain about whether some particular regulation will promote or suppress economic growth.

We can modify the **Philanthropy** choice situation to incorporate this possibility of descriptive uncertainty. For instance, consider the following stylised choice situation:

Risky Philanthropy: some philanthropist is deciding where to donate her fortune. She faces a choice between two options: (1) a charity that provides deworming pills to distant children; and (2) investing in a new social enterprise corporation. Suppose

the philanthropist thinks there is some chance that the social enterprise corporation will benefit a large number of people in her local community. However, she also thinks there is some chance that it will fail to benefit anyone. This philanthropist has 60% credence in an impartialist moral theory T_1 according to which aiding distant strangers is *ceteris paribus* no less choiceworthy than aiding her local community; but she also has 40% credence in a partialist moral theory T_2 according to which aiding her local community is *ceteris paribus* somewhat more choiceworthy than aiding distant strangers.

And as we always do, let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to **Risky Philanthropy**. Moreover, let us also assume that the choiceworthiness returns to scale from funding deworming and soup kitchens are *constant* according to both of the moral theories T_1 and T_2 .

Under this set of circumstances, our philanthropist is descriptively uncertain about how many people will be benefitted if she invests her money in the social enterprise corporation. However, she unfortunately still has to decide how much money to invest in the corporation under these conditions of uncertainty about its success.

In order to handle this descriptive uncertainty, we should stipulate that

in the IMM model for a case like **Risky Philanthropy**, the theory representatives R_1 and R_2 ‘inherit’ their decision maker’s descriptive credences. In particular, R_1 and R_2 will both be uncertain about how many people would be benefitted if the philanthropist invested in the local social enterprise corporation.

At the moment, we are obviously missing several pieces of information that we would need in order to determine which donation distribution would be most appropriate in **Risky Philanthropy** according to IMM. For one thing, we do not yet know exactly what chance of success the philanthropist thinks the social enterprise corporation has; nor do we know exactly how many people will be benefitted if it succeeds.

Moreover, I have also not yet specified T_1 or T_2 ’s views about ethical decision making under the philanthropist’s conditions of descriptive uncertainty. Some possible moral theories will be *risk neutral* – by which I mean that, according to these moral theories, the choiceworthiness of selecting any given risky lottery over possible outcomes must always be equal to a probability-weighted average of what the choiceworthinesses according to this moral theory would be for options to determinately realise each of the possible final outcomes of this risky lottery. However, there are also other possible moral theories which will be *risk averse* – by which I mean that, according to these moral theories, the choiceworthiness of selecting any given risky lottery over possible outcomes must always be *lower* than a probability-weighted average of what the choiceworthinesses according to this moral theory would be for

options to determinately realise each of the possible final outcomes of this risky lottery.¹

If the moral theory T_2 is sufficiently risk averse in the **Risky Philanthropy** choice situation, and if our philanthropist is sufficiently uncertain about the moral value of donating to the social enterprise, then T_2 's representative R_2 might prefer the 'safe option' of our philanthropist donating all of her fortune to deworming (even though according to T_2 , aiding one's local community is *ceteris paribus* more important than aiding distant strangers). Under these conditions, my preferred version of IMM would of course imply that it is most appropriate for the philanthropist to donate all of her fortune to deworming.

By contrast, however, if the moral theory T_2 is risk neutral and our philanthropist thinks that donating to the social enterprise has a strong chance of benefitting a large number of people, then T_2 's representative R_2 will probably prefer for the philanthropist to donate as much of her fortune as possible to the social enterprise corporation. Thus, under these conditions R_1 and R_2 will have diametrically opposed preferences over how the philanthropist should divide her fortune between deworming and the social enterprise. Hence my preferred version of IMM would imply that a 60:40 split between deworming and the social enterprise is the most appropriate dona-

¹For the sake of completeness, I should also mention that there are other possible moral theories which will be *risk loving* – by which I mean that, according to these theories, the choiceworthiness of selecting any given risky lottery over possible outcomes must always be *greater* than our probability-weighted average.

tion distribution in **Risky Philanthropy**.

Chapter 4

Intertemporal dynamics

4.1 Intertemporal bargaining

In every choice situation that I have considered thus far in this dissertation, I have stipulated that our decision maker believes with certainty that she will never afterwards encounter any other choice situations wherein any of the moral theories in which she has credence issue incompatible directives. In each case, I made this stipulation in order to temporarily defer consideration of the complexities introduced by intertemporal gains from trade or contract.

However, almost all morally uncertain decision makers in the real world know that the moral theories in which they have credence will issue incompatible directives in many choice situations that they are likely to confront in the future. For example, although I have credence in some moral theories that will direct me not to bother to vote in the next congressional elections,

I also have credence in some other moral theories according to which voting will be my duty. To take another example, I also know that I will be morally uncertain tomorrow morning about whether or not to order eggs for breakfast.

We can modify the **Double Distribution** set of circumstances (from §2.5 above) to incorporate this possibility of moral uncertainty in future choice situations. For instance, consider the following stylised set of circumstances:

Inheritance: some decision maker is choosing how to divide her free time between two possible uses: one of which is campaigning for nuclear disarmament, and the second of which is volunteering at a local orphanage. Furthermore, this decision maker is also certain that within the next few months, she will inherit some money from her dying grandmother, which she will then have to distribute between two charities: one of which provides deworming pills to distant children, and the second of which supports local soup kitchens. As in **Double Distribution**, suppose that this decision maker has, say, 50% credence in the moral theory T_1 according to which she should spend as much of her free time as possible campaigning for nuclear disarmament, and should eventually donate as much of her inheritance as possible to the deworming charity. She also has 50% credence in the moral theory T_2 according to which she should spend as much of her free time as possible at the local orphanage, and should eventually donate

as much of her inheritance as possible to the local soup kitchens.

And as we always do, let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to **Inheritance**. Moreover, let us also assume that the choiceworthiness returns to scale from each of the possible uses for time and money available in **Inheritance** are *constant* according to both of the moral theories T_1 and T_2 .

Under this set of circumstances, our morally uncertain decision maker now has to choose how to use her free time. Yet she also knows that at some point in the future, she will have to choose what to do with her inheritance. Furthermore, she knows that T_1 and T_2 's directives in both of these two choice situations will be mutually incompatible.

In the IMM model for this **Inheritance** set of circumstances, R_1 and R_2 will each be initially endowed with control rights over 50% of the decision maker's free time. Moreover, R_1 and R_2 should also both be certain that once the decision maker receives her inheritance from her grandmother, these two theory representatives will then each be further endowed with control rights over 50% of that monetary windfall.

In this IMM model for **Inheritance**, R_1 and R_2 should also still have the ability to negotiate trades and contracts with each other. Of course, any trades between R_1 and R_2 will have to be *intertemporal* – in the sense that they will be 'buy now, pay later' kinds of deals. For instance, perhaps R_2

could agree to transfer to R_1 all of R_2 's initial endowment of control rights over the decision maker's free time, in return for R_1 agreeing to transfer to R_2 , all of R_1 's future endowment of control rights over the monetary windfall that our decision maker will eventually inherit from her grandmother.

More generally, I want to stipulate that in the IMM model for any decision maker, the theory representatives should have the ability to negotiate intertemporal trades and contracts potentially ranging over the entire future lifetime of that decision maker. For instance, one of my theory representatives R_2 could agree to transfer to another theory representative R_1 all of R_2 's endowment of control rights over my free time this afternoon, in return for R_1 agreeing to transfer to R_2 , say, 0.1% of R_1 's future endowment of control rights over the monetary windfall that I will receive when I cash out my 401(k) retirement savings account in thirty years time from now.

One could, I suppose, try to imagine an alternative version of IMM according to which bargaining should be allowed to occur only within the boundaries of each particular choice situation confronted by the decision maker. However, any such version of IMM would be implausibly insensitive to intertemporal differences in relative stakes.¹ For example, imagine that I have 50% credence in each of the two moral theories T_1 and T_2 . Suppose that according to T_1 , the stakes are much higher in the choice situation S_1 that I am in right now than they will be in the choice situation S_2 that I am likely

¹These versions of IMM would also face the '*problem of scenario individuation*' (Greaves and Cotton-Barratt 2024, §10) – although cf. Kaczmarek, Lloyd and Plant 2025, §6 for one potential response to this problem.

to encounter sometime next week. By contrast, according to T_2 , the stakes will be much higher in S_2 than they are right now in S_1 . Under these conditions, it is plausible to suppose that it would be most appropriate for me to follow T_1 in S_1 , and T_2 in S_2 . However, IMM can honour this intuition only if it allows for intertemporal bargaining. (Furthermore, I will argue in future work that allowing for intertemporal bargaining also ensures that IMM can honour our intuitions about the value of moral inquiry in response to moral uncertainty.)

Of course, allowing for intertemporal bargaining over one's entire future lifetime will introduce a large amount of descriptive uncertainty into the IMM models for almost all real-world decision makers. After all, few of us know exactly which choice situations we will confront a few years from now – let alone a few decades! We can slightly modify the **Inheritance** example to incorporate this element of descriptive uncertainty. For instance, consider the following stylised set of circumstances:

Risky Inheritance: some decision maker is choosing how to divide her free time between disarmament campaigning and orphanage volunteering. Furthermore, this decision maker is also certain that within the next few months, she will inherit some money from her dying grandmother, which she will then have to distribute between deworming and soup kitchens. However, this decision maker is uncertain about how much money her grandmother will bequeath to her. Suppose that this decision maker

has 99% credence in T_1 , which favours disarmament campaigning and deworming; with 1% credence in T_2 , which favours orphanage volunteering and soup kitchens.

(It will become clear in §4.3 below why I have switched the decision maker's moral credence distribution here to 99:1.) As we always do, let us also assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to **Risky Inheritance**. And let us also continue to assume constant returns to scale.

In the IMM model for this **Risky Inheritance** set of circumstances, R_1 will be initially endowed with control over 99% of the decision maker's free time, and R_2 will be initially endowed with control over the remaining 1%. Moreover, once our decision maker receives her inheritance, R_1 and R_2 will be respectively endowed with control rights over 99% and 1% of this monetary windfall. Since the decision maker is descriptively uncertain about how much money she will inherit from her grandmother, we should also stipulate that R_1 and R_2 both share the decision maker's descriptive credences over the size of her inheritance (as in §3 above).

In this model for **Risky Inheritance**, R_1 and R_2 should once again have the ability to negotiate intertemporal trade and contracts with each other. However, in light of our decision maker's uncertainty about how much she will inherit, R_1 and R_2 will be uncertain about exactly how happy *ex post*

each of these two theory representatives would be with a contract under which – for instance – R_2 agrees to transfer to R_1 all of R_2 's endowment of control rights over free time, in return for R_1 agreeing to transfer to R_2 , say, 2% of R_1 's future endowment of control rights over however much money the decision maker eventually inherits from her grandmother.

Although a contract like this would thus involve an element of risk, agreeing to this kind of contract might nonetheless be in the best interests of both R_1 and R_2 . If R_1 cares about free time much more than she cares about money, then she might benefit from this kind of contract as a way of gaining greater control over the decision maker's free time. And if R_2 cares much more about money than she cares about free time, then she might benefit from this kind of contract as a way of gaining greater control over however much money the decision maker eventually inherits from her grandmother.

4.2 Noncompliance

One new question raised by the possibility of intertemporal bargaining concerns cases wherein our decision maker has behaved inappropriately at some earlier point in time, by failing to conform to the instructions that the theory representatives in her IMM model would have jointly issued to her. In such cases, should the decision maker's past inappropriate behaviour alter how it would be most appropriate for this decision maker to behave going forward?

For example, imagine that in the **Inheritance** set of circumstances, T_1

implies that how the decision maker uses her free time is much more important than how she will use her inheritance, whereas T_2 implies that how the decision maker will use her inheritance is much more important than how she uses her free time. Furthermore, imagine that R_1 and R_2 's bargaining positions in **Inheritance** are structurally identical to each other, in the sense defined in §2.4 above. Thus (recalling §2.5 above), we can safely assume that R_2 will agree to transfer to R_1 all of R_2 's initial endowment of control over free time, in return for R_1 agreeing to transfer to R_2 all of R_1 's initial endowment of control over the money that our decision maker will inherit from her grandmother. After this trade, R_1 will instruct our decision maker to spend all of her free time campaigning for nuclear disarmament, whereas R_2 will instruct her to donate all of her inheritance to soup kitchens.

However, now imagine that our decision maker does not think through – or else does not care about – any of these recommendations from IMM in deciding what to do with her free time. Hence, suppose that our decision maker inappropriately chooses to spend all of her free time at the local orphanage. Then, later on, she eventually receives her inheritance. Under these circumstances, how would it be most appropriate for our decision maker to now distribute her inheritance between deworming and soup kitchens? Let's call this the *problem of noncompliance*.²

²In this dissertation, I will focus on cases in which our decision maker knows that she has acted inappropriately in the past. However, one can also imagine cases in which our decision maker predicts that she *will* act inappropriately at some time in the *future*. For instance, we could construct a case like this inspired by the 'Professor Procrastinate' case, which has been discussed at great length in philosophical debate between ethical actualists

4.2.1 Responses

I will now sketch three possible responses to this problem of noncompliance.³

(1) **NONCONTAMINATIONISM**: According to the **NONCONTAMINATIONIST** response, our decision maker having previously used her free time inappropriately should in no way alter which donation distribution would be most appropriate once she receives her inheritance. Hence, according to **NONCONTAMINATIONISM**, it is still most appropriate for our decision maker to donate all of her inheritance to soup kitchens – regardless of how she previously distributed her free time. More generally, **NONCONTAMINATIONISM** is the view that past noncompliance never alters whether it is appropriate for our decision maker to follow the original intertemporal contracts negotiated by her theory representatives (at least in cases like **Inheritance** wherein past noncompliance would not alter the marginal choiceworthinesses of the options available in future).

(2) **NULLIFICATIONISM**: According to the **NULLIFICATIONIST** re-

and possibilists. (The original version of the case is from Goldman 1978. For an overview of the broader debate, see Timmerman and Cohen 2019.) Although I will not discuss future noncompliance in this dissertation, all three of the responses to past noncompliance that I present in §4.2.1 below could be naturally extended to cover future noncompliance cases too.

³Assume, for the rest of this section, that our decision maker’s credence distribution over moral theories never changes over the course of her lifetime. I will relax this assumption in §4.4 below.

sponse, our decision maker having previously used her free time inappropriately *voids* the intertemporal contract originally negotiated between R_1 and R_2 . Hence, according to NULLIFICATIONISM, once our decision maker receives her inheritance, R_1 and R_2 should each be endowed with control rights over 50% of it, with neither theory representative being under any contractual obligation to transfer control rights to the other. After this inheritance endowment, R_1 and R_2 will then have the ability to negotiate new trades or contracts with each other (from a clean slate) – although there will in fact be no gains from trade or contract available under these conditions after our decision maker inherits. More generally, NULLIFICATIONISM is the view that whenever a decision maker acts inappropriately, this should be taken to nullify any intertemporal contracts pertaining to this choice that were previously negotiated by her representatives.

(3) COMPENSATIONISM: According to the COMPENSATIONIST response, although our decision maker having inappropriately used her free time in the manner favoured by T_2 rather than T_1 does not *void* R_1 and R_2 's intertemporal contract, it does mean that the execution of this contract should be immediately followed by R_2 'compensating' R_1 through a supplementary transfer of control rights over some of the decision maker's inheritance. If R_2 owed R_1 only a small amount of compensation under these conditions,

then R_2 could just be required to *return* to R_1 some fraction of the control rights that R_1 was first of all contractually obligated to transfer to R_2 . However, if R_2 owed R_1 a very large amount of compensation under our conditions, then R_2 could be required to transfer to R_1 *all* of R_2 's holding of control rights over the inheritance, thereby giving R_2 total control over how the decision maker will be instructed to spend her monetary windfall. In general, COMPENSATIONISM is the view that when a decision maker inappropriately chooses an option that is dispreferred over the appropriate option by certain moral theories in which the decision maker has credence, then – *ceteris paribus* – the representatives of those moral theories are afterwards *pro tanto* entitled to some compensation. Different possible precisifications of COMPENSATIONISM will use different possible principles to determine exactly how much compensation should be transferred to the affected representatives.

These three possible responses to the problem of noncompliance all have different implications for which donation distribution would be most appropriate in **Inheritance** after our decision maker has inappropriately spent all of her free time at the local orphanage (as favoured by T_2). NONCONTAMINATIONISM implies that even after this noncompliance, it would still be most appropriate for the decision maker to donate all of her inheritance to soup kitchens (again, as favoured by T_2). By contrast, NULLIFICATION-

ISM implies that after this noncompliance, it is most appropriate to split the inheritance 50:50. Finally, COMPENSATIONISM implies that the appropriate donation share for deworming must be nonzero, with the appropriate donation share for soup kitchens correspondingly being less than 100%. (Different precisifications of COMPENSATIONISM will imply different appropriate donation shares within these constraints.)

Although NULLIFICATIONISM is initially at least somewhat intuitively appealing, its implications are much less plausible than those of NONCONTAMINATIONISM and COMPENSATIONISM in certain ‘near miss’ noncompliance cases. For example, imagine that in some modified version of **Inheritance**, our decision maker also has the option to spend some or all of her free time campaigning for decarbonization. Suppose that decarbonization campaigning is very slightly less choiceworthy than disarmament campaigning according to T_1 , but very slightly more choiceworthy than this according to T_2 . Hence, decarbonization campaigning is much more choiceworthy than orphanage volunteering according to T_1 , but is much less choiceworthy than this according to T_2 (as illustrated in figure 4.1).

Suppose that the most appropriate course of action in this modified version of **Inheritance** would still be for our decision maker to spend all of her free time on disarmament campaigning, thereafter donating all of her inheritance to soup kitchens. However, now imagine that our decision maker in fact inappropriately chooses to spend all of her free time on decarbonization campaigning. Then, later on, she eventually receives her inheritance

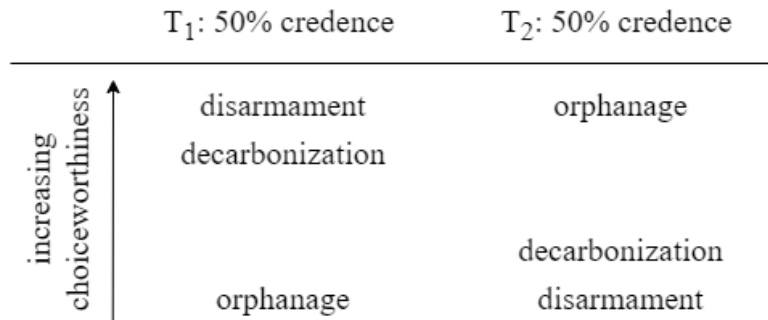


Figure 4.1: Uses for free time in our modified version of **Inheritance**

windfall.

Under these noncompliance conditions, NONCONTAMINATIONISM implies that it would still be most appropriate for our decision maker to donate all of her inheritance to soup kitchens. Furthermore, any plausible version of COMPENSATIONISM would imply that it will be most appropriate to donate only a very small percentage of the inheritance to deworming, and hence only slightly less than 100% to soup kitchens. After all, decarbonization campaigning is only very slightly less choiceworthy than disarmament campaigning according to R_1 , and only very slightly more choiceworthy than this according to R_2 . Hence, any plausible version of COMPENSATIONISM will imply that if our decision maker inappropriately campaigns for decarbonization instead of disarmament, then R_2 should afterwards owe to R_1 only a very limited transfer of compensation.

By contrast, however, under these noncompliance conditions, NULLIFICATIONISM implies that it would be most appropriate for our decision maker to

split her inheritance 50:50 between deworming and soup kitchens. After all, our decision maker acted inappropriately in choosing to spend her free time campaigning for decarbonization rather than disarmament. Hence, NULLIFICATIONISM implies that once our decision maker receives her inheritance, R_1 and R_2 should each be endowed with control rights over 50% of it, with neither theory representative being under any contractual obligation to transfer any control rights to the other. But under these conditions, R_1 will simply instruct our decision maker to donate 50% of her inheritance to deworming, whereas R_2 will instruct her to donate 50% to soup kitchens.

Unfortunately for NULLIFICATIONISM, it strikes me as highly implausible to suppose that it is most appropriate for our decision maker to split her inheritance 50:50 between deworming and soup kitchens (under these conditions of noncompliance). T_1 and T_2 both imply that there is only a very small difference in choiceworthiness between campaigning for disarmament versus decarbonization. Hence, it is implausible to suppose that our decision maker campaigning for decarbonization as opposed to disarmament could shift the most appropriate inheritance distribution between deworming and soup kitchens from 0:100 to 50:50. This goes to show that NULLIFICATIONISM has implausible implications under ‘near miss’ noncompliance conditions.

With NULLIFICATIONISM off the table, we now face a choice between NON-CONTAMINATIONISM and COMPENSATIONISM. Each of these two positions suggests a certain underlying view about how the appropriateness of any particular choice or option should depend upon IMM’s basic desideratum of

affording each moral theory its due influence.

Firstly, COMPENSATIONISM suggests that the appropriateness of any particular option should depend upon the the extent to which choosing that option would bring each moral theory closer to having had its due influence over the decision maker's total lifetime course of action. According to COMPENSATIONISM, if our decision maker previously failed to give some moral theory its due influence over her choices, then it would be most appropriate for her to now correct this past imbalance by temporarily giving that moral theory greater influence over her decisions than would have been appropriate otherwise. For instance, if our decision maker previously failed to give T_1 due influence over how she spent her free time in **Inheritance**, then it would be most appropriate for her to now correct this past imbalance by giving T_1 greater influence over how she distributes her inheritance than would have been appropriate otherwise.

On the other hand, NONCONTAMINATIONISM suggests that the appropriateness of any particular option should depend only upon whether this option is part of a total lifetime 'IDEAL PLAN' of action that would come as close as possible to giving each moral theory its lifetime due influence. According to NONCONTAMINATIONISM, even if our decision maker has previously failed to follow her lifetime IDEAL PLAN of action, it would still be most appropriate for her to now follow that IDEAL PLAN of action as closely as possible. For instance, even if our decision maker previously used her free time inappropriately in **Inheritance**, it would still be most appropriate for her to donate

all of her inheritance soup kitchens.

The difference between these two possible views about appropriateness is roughly analogous to the difference between the ‘*act*’ and ‘*rule*’ versions of utilitarianism. Act utilitarianism says that any given action is permissible iff that action would maximise total wellbeing. In other words, an action’s permissibility should *directly* depend upon the desideratum of maximising wellbeing. On the other hand, rule utilitarianism says that any given action is permissible iff that action would be recommended by the total code of rules that would maximise total wellbeing. In other words, an action’s permissibility should directly depend upon the ideal code of rules, and thus should depend only *indirectly* upon the desideratum of maximising wellbeing.

Analogously, the COMPENSATIONIST view says that an option’s appropriateness should *directly* depend upon the desideratum of due influence. By contrast, the NONCONTAMINATIONIST view says that an option’s appropriateness should directly depend upon the lifetime IDEAL PLAN of action, and hence should depend only *indirectly* upon the desideratum of due influence.

So understood, COMPENSATIONISM strikes me as a more attractive view than NONCONTAMINATIONISM. If overall due influence is really a legitimate desideratum, then it seems to me most plausible to suppose that the appropriateness of any given option should *directly* depend upon how this option stands with respect to the due influence desideratum. At the very least, NONCONTAMINATIONISTS owe us an explanation of why an option’s appropriateness should depend only indirectly upon the desideratum of due influence.

The analogy between the intrapersonal problem of moral uncertainty and the interpersonal problem of incompatible desires that I used to motivate IMM in §1.3 above might also be taken to provide some support for COMPENSATIONISM. In the social setting, it strikes me as plausible to suppose that whenever one party to a contract receives less than she bargained for, she should thereafter be entitled to some compensation. And something like this principle is of course also reflected in how our real-world legal systems handle breaches of contract.

However, we should be wary of placing too much weight on this analogical argument in favour of COMPENSATIONISM. I have argued (in §1.3 above) that the intrapersonal problem of moral uncertainty is to some extent analogous to the interpersonal problem of incompatible desires, and that this analogy motivates IMM's core ideas. However, I do not take myself to have shown that the analogy between these two problems is watertight enough that it should serve as a *straightjacket* on our development of all the details of IMM. Hence, the argument from analogy should provide only limited and defeasible evidence in favour of COMPENSATIONISM.

Still, my own view is that we should adopt the COMPENSATIONIST response to the problem of noncompliance. However, I will not try to settle the debate between NONCONTAMINATIONISM and COMPENSATIONISM in this dissertation. Instead, I will simply leave this as an unsettled theoretical choice point.

4.2.2 Compensation

As I have already mentioned, COMPENSATIONISM can be combined with various different principles for determining exactly what compensatory transfers should be required after any given instance of noncompliance. However, in making our choice among these various possible compensation principles, I suggest that we should be guided by the motivating view that the appropriateness of any particular option ought to depend upon the extent to which choosing that option would bring each moral theory closer to getting at least its due influence over the decision maker's total lifetime course of action.

Hence, our noncompliance compensation principle should always come as close as possible to giving each theory representative at least the level of influence on our decision maker's overall course of behaviour that this representative would have enjoyed had our decision maker always afforded each moral theory its due influence. Moreover, I want to suggest that

in response to any given noncompliance set of circumstances, a COMPENSATIONIST version of IMM would give every theory representative its due influence IFF this version of IMM would render every theory representative *indifferent* between (a) what the decision maker's total lifetime course of behaviour *would have* been had she *always* followed the IDEAL PLAN instructions of her IMM theory representatives, and (b) what the decision maker's 'PARTIALLY NONCOMPLIANT' total lifetime course of behaviour

will in fact be, supposing only that she will follow the instructions of her IMM theory representatives in all *future* choice situations.

Call this the INDIFFERENCE CRITERION.

For example, if the decision maker in **Inheritance** inappropriately chose to spend all of her free time at the local orphanage (as favoured by R_2), then the INDIFFERENCE CRITERION implies that a COMPENSATIONIST version of IMM would give every theory representative its due influence IFF this version of IMM would render every theory representative indifferent between

(a) the IDEAL PLAN overall course of behaviour wherein our decision maker (a1) spent all of her free time campaigning for nuclear disarmament (as favoured by R_1), and (a2) will donate all of her inheritance to soup kitchens (as favoured by R_2)

and

(b) the PARTIALLY NONCOMPLIANT overall course of behaviour wherein our decision maker (b1) spent all of her free time volunteering at the local orphanage (as favoured by R_2), but (b2) will split her inheritance in the manner recommended by this COMPENSATIONIST version of IMM.

After all, if any given theory representative preferred the IDEAL PLAN over PARTIAL NONCOMPLIANCE, then PARTIAL NONCOMPLIANCE would plausibly give this theory representative *less* than its due influence over our decision

maker's overall response to **Inheritance**. But on the other hand, if any given theory representative preferred PARTIAL NONCOMPLIANCE over the IDEAL PLAN, then PARTIAL NONCOMPLIANCE would plausibly give this theory representative *more* than its full due influence over the decision maker's behaviour. Thus, the only way for *every* theory representative to have its due influence under these circumstances of noncompliance is for every representative to be indifferent between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour.

Unfortunately, however, we can safely assume that it will be impossible for any COMPENSATIONIST version of IMM to render every theory representative indifferent between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour under these conditions. Because R_1 cares about how the decision maker uses her free time much *more* than about how the decision maker uses her inheritance, R_1 could be rendered indifferent between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour only by a version of IMM that required R_2 to transfer to R_1 control over a relatively *large* fraction of our decision maker's inheritance. However, R_2 by contrast cares about how the decision maker uses her free time much *less* than about how the decision maker uses her inheritance; and so R_2 could be rendered indifferent between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour only by a version of IMM that required R_2 to transfer to R_1 control over a relatively *small* fraction of our decision maker's inheritance. Thus, no possible scheme of compensation transfers could render both R_1 and R_2 indifferent between

the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour, since no fraction of our decision maker's inheritance can simultaneously be both 'large' and 'small.'⁴

Which scheme of transfers should our COMPENSATIONIST principle select in these noncompliance circumstances, wherein it is impossible to perfectly satisfy the INDIFFERENCE CRITERION? Well, in any set of circumstances like this, it strikes me as *prima facie* plausible to suppose that our COMPENSATIONIST principle should select the scheme of compensation transfers that would *minimise the shortfall* from giving all of the theory representatives at least their due influence. In other words,

in response to any given noncompliance set of circumstances, we should select a scheme of compensation transfers that minimises the shortfall from rendering every theory representative at worst indifferent between (a) what the decision maker's total lifetime course of behaviour would have been had she always followed the 'IDEAL PLAN' instructions of her IMM theory representatives – call this ' χ_{ideal} ' for short, and (b) what the decision maker's 'PARTIALLY NONCOMPLIANT' total lifetime course of behaviour

⁴In fact, we can (much more generally) safely assume that it will *always* be impossible for any COMPENSATIONIST version of IMM to render every theory representative indifferent between the IDEAL PLAN and any possible PARTIALLY NONCOMPLIANT course of behaviour. After all, surely our decision maker can have acted inappropriately only if this decision maker has done something that is inconsistent with every theory representative having at least its lifetime due influence. Thus, if our decision maker has acted inappropriately, then it must now be impossible to render every theory representative indifferent between the IDEAL PLAN and any possible completion of the PARTIALLY NONCOMPLIANT course of behaviour.

will in fact be, supposing only that she will follow the instructions of her IMM theory representatives in all future choice situations – call this ‘ χ_{pnc} ’ for short.

Call this the SHORTFALL CRITERION. Taken together with COMPENSATIONISM, this SHORTFALL CRITERION implies that:

the most appropriate course of action is response to any given noncompliance set of circumstances is the course of action ϕ which would minimise the shortfall from rendering every theory representative at worst indifferent between (a) what the decision maker’s lifetime ‘IDEAL PLAN’ course of behaviour would have been had she always followed the instructions of her IMM representatives (χ_{ideal}), and (b) what the decision maker’s ‘PARTIALLY NONCOMPLIANT’ total lifetime course of behaviour will in fact be, supposing only that she will follow ϕ in all future choice situations (χ_{pnc}).

In order to precisify the SHORTFALL CRITERION, we will need to specify both: (1) how to measure the extent to which any given scheme of compensation transfers would ‘fall short of’ rendering any particular theory representative at worst indifferent between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour; and also (2) how to *aggregate* these measurements for each individual theory representative into an overall measure of the extent to which our scheme of compensation transfers would ‘fall short of’ rendering

every theory representative at worst indifferent between these two possible courses of behaviour. I will discuss these two questions in §§4.2.3–4.2.4 below.

4.2.3 Individual shortfall

At first, it might seem natural to measure the extent to which any given theory representative R_i falls short of indifference between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour in terms of the *absolute choiceworthiness difference* $CW_i(\mathbf{x}_{\text{ideal}}) - CW_i(\mathbf{x}_{\text{pnc}})$ between them according to this theory representative. Unfortunately, however, these absolute choiceworthiness measurements can be aggregated into an overall measurement of shortfall only if we can make intertheoretic choiceworthiness comparisons. And as I have already noted in §1.2 above, it is open to question whether any such intertheoretic choiceworthiness comparisons are possible. In fact, I myself think that intertheoretic choiceworthiness comparisons are always impossible. Thus, my own view is that we should not measure shortfall in terms of absolute choiceworthiness differences.

How else can we measure the extent to which any given theory representative falls short of indifference between the ideal and the partially non-compliant courses of behaviour? Unfortunately, I will only be able to give a completely precise statement of own preferred approach after I have introduced the necessary formal apparatus in §7.3 below. In the meantime, however, I can at least give a rough explanation of the main idea.

A first-pass statement of my main idea here is that we should measure the

extent to which any given theory representative falls short of having at least its due influence in terms of the extent to which this theory representative is closer to having its due influence than it is to having exactly *zero* influence. Thus, I suggest that the extent to which the partially noncompliant course of behaviour χ_{pnc} falls short of giving any particular theory representative R_i its due influence should be measured *relative* to the choiceworthiness difference according to R_i between (1) the ideal course of behaviour, and (2) the 'IDEAL-WITHOUT- R_i ' course of behaviour ' χ_{-i} ' that our decision maker would have been instructed to choose by our IMM model if R_i had not been endowed with any control rights, and hence had zero influence over our decision maker's behaviour. In other words, we should measure the extent to which R_i falls short of indifference between χ_{ideal} and χ_{pnc} by something like the relative choiceworthiness difference

$$\frac{CW_i(\chi_{\text{ideal}}) - CW_i(\chi_{\text{pnc}})}{CW_i(\chi_{\text{ideal}}) - CW_i(\chi_{-i})}$$

where:

IF c denotes our decision maker's credence function over moral theories,

THEN χ_{-i} can be defined as the course of behaviour that our IMM model would have instructed our decision maker to choose *if* her credence function over moral theories had instead been c_{-i} ,

where⁵

$$c_{-i}(t) := \begin{cases} 0 & \text{if } t = T_i \\ \frac{c(t)}{1-c(T_i)} & \text{otherwise} \end{cases}$$

Notice that this new formula measures the choiceworthiness difference between χ_{ideal} and χ_{pnc} as a *percentage* of the choiceworthiness difference between χ_{ideal} and χ_{-i} . Thus, this measure of distance from indifference is defined in terms of relative rather than absolute choiceworthiness. Hence, even if intertheoretic unit comparisons are impossible, it would still make perfect sense for us to intertheoretically compare and aggregate this new relative measure of distance from indifference. This kind of aggregation would only require us to say things like: ‘moving 20% closer to due than to zero influence is a relatively greater improvement than moving only 10% closer to due than to zero influence.’ Whereas, by contrast, aggregating a measure defined in terms of absolute choiceworthinesses would instead require intertheoretic comparisons, of the form: ‘moving 5 T_1 -units of choiceworthiness closer to due influence is an absolutely greater improvement than moving only 3 T_2 -units of choiceworthiness closer to due influence.’ Hence, my new approach to measuring shortfall from indifference can be aggregated into an overall shortfall measure even if intertheoretic unit comparisons are impossible.⁶

⁵For those unfamiliar with this notation, ‘:=’ abbreviates ‘is defined as being equal to.’

⁶That being said, my first-pass statement of this relative-choiceworthiness formula cannot be used with any moral theories according to which we cannot use an *interval scale* of choiceworthiness to measure the strengths of our decision maker’s all-things-considered

I will discuss my preferred approach to measuring shortfalls from indifference in greater detail in §7.3 below, and so for now I will simply focus on characterizing how this proposal works in my running example of noncompliance for the **Inheritance** set of circumstances. Recall that in this example, our decision maker has 50% credence in each of the two moral theories T_1 and T_2 . According to the moral theory T_1 , our decision maker should spend as much free time as possible campaigning for nuclear disarmament, and she should donate as much of her inheritance as possible to deworming. Hence, the IDEAL-WITHOUT- R_2 response to **Inheritance** which would result from R_1 having total control over our decision maker's behaviour would require this decision maker to spend all of her time campaigning for disarmament, and for her to donate all of her inheritance to deworming. By contrast, according to the moral theory T_2 , our decision maker should spend as much free time as possible volunteering at a local orphanage, and she should donate as much of her inheritance as possible to local soup kitchens. Hence, the IDEAL-WITHOUT- R_1 response to **Inheritance** which would result from R_2 having total control over our decision maker's behaviour would require this decision maker to spend all of her time volunteering at the orphanage, and for her to donate all of her inheritance to soup kitchens. These results are summarized in figure 4.2.

I have also stipulated that in **Inheritance**, T_1 implies that how our decision maker uses her free time is more important than how she will use her

moral reasons. I address this problem in §7.3 below.

resource:	time	money
R_1 's preferred use:	nuclear disarmament	deworming
R_2 's preferred use:	orphanage	soup kitchens
representative that cares about this resource more than the other resource:	R_1	R_2
IDEAL PLAN:	nuclear disarmament	soup kitchens
PARTIAL NONCOMPLIANCE:	orphanage	?
IDEAL-WITHOUT- R_1 :	orphanage	soup kitchens
IDEAL-WITHOUT- R_2 :	nuclear disarmament	deworming

Figure 4.2: Courses of behaviour in **Inheritance**

inheritance, whereas T_2 has the reverse implication. In fact, I assumed that R_1 and R_2 's bargaining positions are structurally identical to each other. Hence, the IDEAL PLAN response to **Inheritance** according to IMM would require our decision maker to spend all of her free time campaigning for nuclear disarmament, and then after that to donate all of her inheritance to the soup kitchens. But of course, in my running example of *noncompliance* in **Inheritance**, I have been assuming that our decision maker in fact inappropriately chooses to spend all of her free time at the local orphanage. (Once again, all of this is summarized in figure 4.2.)

Under these conditions, we can now evaluate how close various possible completions of our decision maker's PARTIALLY NONCOMPLIANT course of behaviour would come to giving each of the two theory representatives R_1 and R_2 at least their due influences over our decision maker's total lifetime course of action. Consider, first of all, the possible completion of the PARTIALLY NONCOMPLIANT course of behaviour under which our decision maker would be required to donate all of her inheritance to soup kitchens. This

‘ORPHANAGE-THEN-SOUP KITCHENS’ completion of the PARTIALLY NON-COMPLIANT course of behaviour corresponds to the NONCONTAMINATION-IST (and so *non*-COMPENSATIONIST) response to past noncompliance. I will now confirm that this course of behaviour cannot satisfy the SHORTFALL CRITERION for appropriate compensation transfers.

According to R_1 , ORPHANAGE-THEN-SOUP KITCHENS would be much less choiceworthy than the IDEAL PLAN. In fact, according to R_1 , the choiceworthiness difference between the IDEAL PLAN and ORPHANAGE-THEN-SOUP KITCHENS will obviously be equal to the choiceworthiness difference between the IDEAL and the IDEAL-WITHOUT- R_1 courses of behaviour, since ORPHANAGE-THEN-SOUP KITCHENS is simply *identical* to the IDEAL-WITHOUT- R_1 course of behaviour (recall figure 4.2 above). Thus, my preferred formula for measuring shortfalls from indifference will imply that R_1 falls far short of indifference between the IDEAL PLAN and ORPHANAGE-THEN-SOUP KITCHENS.

On the other hand, according to R_2 , ORPHANAGE-THEN-SOUP KITCHENS will be more choiceworthy than the IDEAL PLAN. In fact, according to R_2 , ORPHANAGE-THEN-SOUP KITCHENS is the most choiceworthy course of behaviour available in **Inheritance**. Hence, ORPHANAGE-THEN-SOUP KITCHENS would certainly give R_2 at least due influence over our decision maker’s overall course of action.

It follows from these results that ORPHANAGE-THEN-SOUP KITCHENS cannot satisfy the SHORTFALL CRITERION for appropriate compensation. To

see why this result obtains, suppose that we compare ORPHANAGE-THEN-SOUP KITCHENS against some alternative completion of the PARTIALLY NON-COMPLIANT course of behaviour under which our decision maker would be required to donate some nonzero $\varepsilon\%$ of her inheritance to deworming (and hence to donate the remaining $(100 - \varepsilon)\%$ to soup kitchens). Let us call this the ‘ORPHANAGE-THEN- $D(\varepsilon)$ ’ course of behaviour. Clearly, for any nonzero value of ε , this new ORPHANAGE-THEN- $D(\varepsilon)$ completion must be more choiceworthy than ORPHANAGE-THEN-SOUP KITCHENS according to R_1 . Thus, ORPHANAGE-THEN- $D(\varepsilon)$ must come closer than ORPHANAGE-THEN-SOUP KITCHENS does to giving R_1 at least due influence.

Furthermore, since ORPHANAGE-THEN-SOUP KITCHENS is *more* choiceworthy than the IDEAL PLAN according to R_2 , there must therefore be some small yet nonzero values of ε for which ORPHANAGE-THEN- $D(\varepsilon)$ would be at least as choiceworthy as the IDEAL PLAN according to R_2 . Hence, for these values of ε , ORPHANAGE-THEN- $D(\varepsilon)$ would still give R_2 at least its due level of influence.

Thus, for at least some nonzero values of ε , ORPHANAGE-THEN- $D(\varepsilon)$ must weakly dominate ORPHANAGE-THEN-SOUP KITCHENS with respect to the desideratum of giving each of our theory representatives at least their due influence. Therefore, ORPHANAGE-THEN-SOUP KITCHENS cannot satisfy the SHORTFALL CRITERION, and so cannot be an appropriate response to this set of noncompliance circumstances according to COMPENSATIONISM.

Having evaluated ORPHANAGE-THEN-SOUP KITCHENS, I will now turn

my attention to what is in some sense the ‘opposite’ completion of PARTIAL NONCOMPLIANCE – *viz.* the ‘ORPHANAGE-THEN-DEWORMING’ course of behaviour. This course of behaviour would require our decision maker to donate all of her inheritance to deworming, as favoured by R_1 . Hence, this completion of PARTIAL NONCOMPLIANCE would correspond to R_1 receiving the maximum possible compensation transfer from R_2 under these conditions of noncompliance.

Given our assumption that R_1 and R_2 have identical bargaining positions in **Inheritance**, these two theory representatives must both fall equally far short of indifference between the IDEAL PLAN and ORPHANAGE-THEN-DEWORMING. After all, each of these two theory representatives regards ORPHANAGE-THEN-DEWORMING as the course of behaviour under which (1) the resource that this representative cares less about is used exclusively as this representative would prefer, although (2) the resource that this representative cares more about is used exclusively as this representative would disprefer. Thus, both of these two theory representatives will regard ORPHANAGE-THEN-DEWORMING as being somewhat less choiceworthy than our decision maker’s IDEAL PLAN.

The exact choiceworthiness of ORPHANAGE-THEN-DEWORMING according to each of these two theory representatives will depend upon the extent to which each of these representatives cares about one of the two resources more strongly than they care about the other. If each of these two theory representatives cares only very slightly more about one or the other of our two

resources, then both of these two representatives will regard ORPHANAGE-THEN-DEWORMING as being very near to the choiceworthiness of the IDEAL PLAN. But, at the other extreme, if each of these two theory representatives cares very much more about one or the other of our two resources, then both of these two representatives will regard ORPHANAGE-THEN-DEWORMING as being much less choiceworthy than the IDEAL PLAN.

Could ORPHANAGE-THEN-DEWORMING satisfy our SHORTFALL CRITERION for appropriate compensation? In order to answer this question, we will need to compare R_1 and R_2 's shortfalls from at least due influence under ORPHANAGE-THEN-DEWORMING against these two representatives' shortfalls under all of the various possible completions of our PARTIALLY NONCOMPLIANT course of behaviour. In other words, we will need to evaluate how R_1 and R_2 's shortfalls from at least due influence will vary as a function of how much of our decision maker's inheritance is donated to deworming after her earlier noncompliance.

We can begin by graphing everything that we know so far. In figure 4.3, the horizontal axis corresponds to the percentage of our decision maker's inheritance that is donated to deworming, and the vertical axis corresponds to shortfall from at least due influence (as measured by my preferred formula). I have already demonstrated that if 0% of our decision maker's inheritance is donated to deworming, then R_1 will have a 100% shortfall from due influence, whereas R_2 will by contrast have greater than her due influence. These results are represented by two dots on the left hand side of figure 4.3.

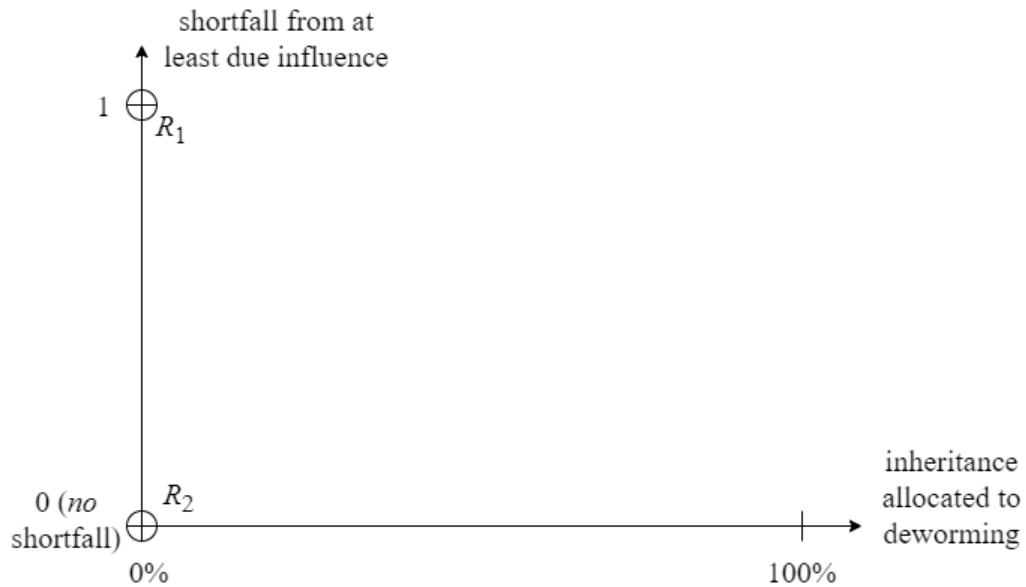


Figure 4.3: Shortfalls when 0% is allocated to deworming

I have also already demonstrated that if 100% of our decision maker's inheritance is donated to deworming, then R_1 and R_2 will both have exactly the same shortfall from due influence. Moreover, if R_1 and R_2 each care roughly the same about time and money, then both of these two theory representatives would have very close to their due level of influence if 100% were donated to deworming – as illustrated in figure 4.4. On the other hand, if R_1 and R_2 each cares much more about one of our two resources than they care about the other, then both R_1 and R_2 would fall far short of their due level of influence if 100% were donated to deworming – as illustrated in figure 4.5.

More generally, we can now try to graph how R_1 and R_2 's shortfalls

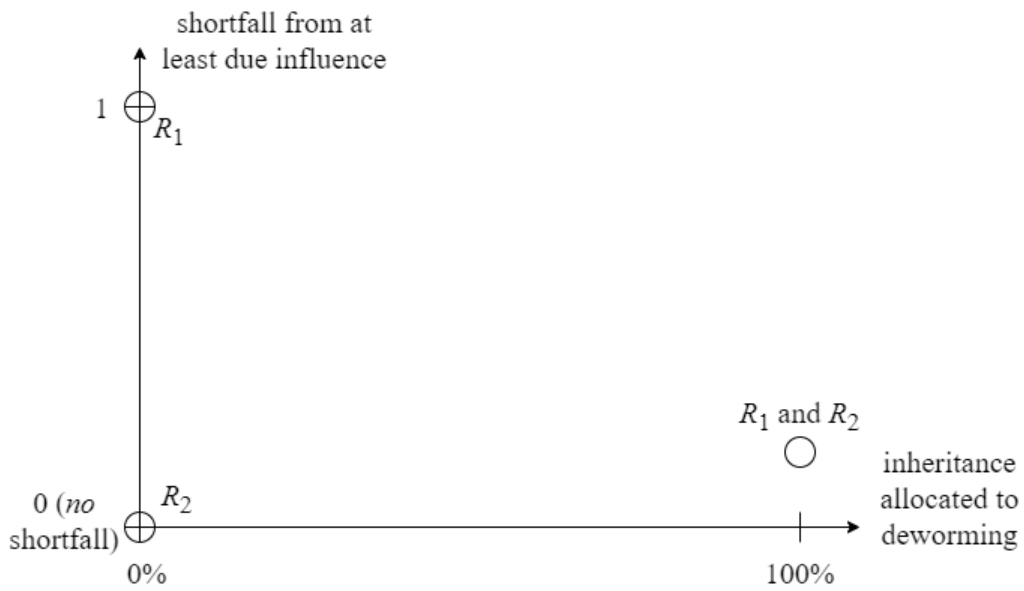


Figure 4.4: Shortfalls if each representative cares roughly the same about time and money

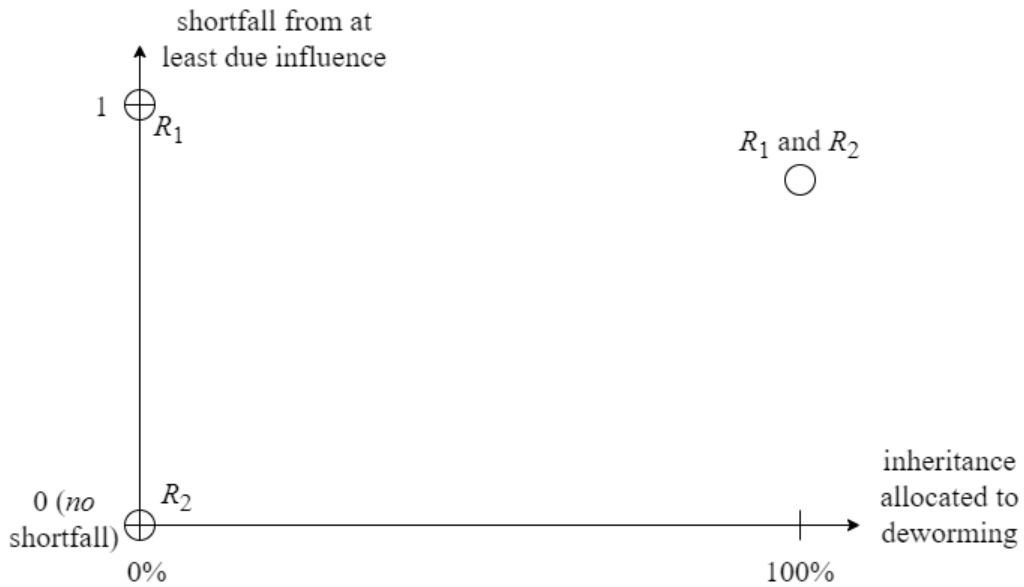


Figure 4.5: Shortfalls if each representative cares much more about one of our two resources

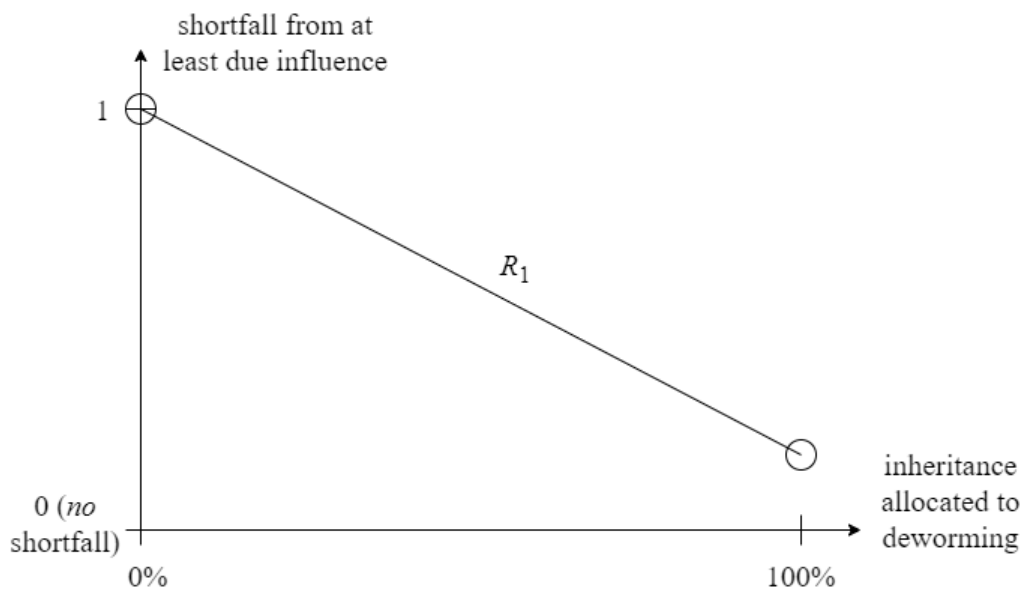


Figure 4.6: R_1 's shortfall function if each representative cares roughly the same about time and money

from at least due influence would vary as a function of the percentage of our decision maker's inheritance that is allocated to deworming. Granted the assumption of constant returns scale, it follows almost immediately that R_1 's shortfall function must be the *straight line* which joins R_1 's two shortfall points at 0 and 100%. For instance, if 50% of the inheritance were to be allocated to deworming, then R_1 's shortfall from due influence would be equal to the mean of her two shortfalls at the 0 and 100% allocations. We can hence update figures 4.4 and 4.5 with a full shortfall function for R_1 – see figures 4.6 and 4.7.

Things are a little more complicated for the representative R_2 's shortfall function. I have already argued that there must be some low yet nonzero

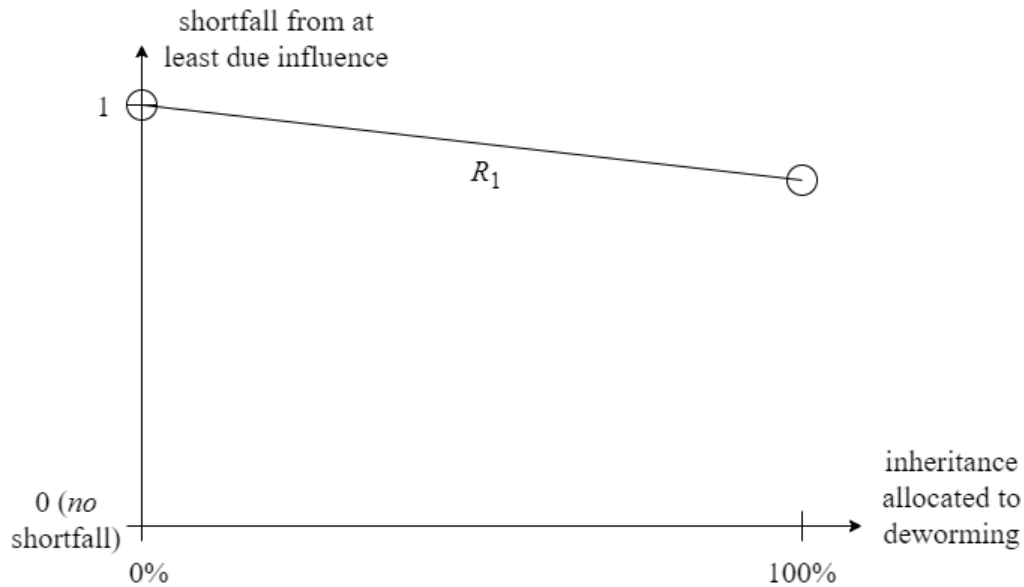


Figure 4.7: R_1 's shortfall function if each representative cares much more about one of our two resources

allocations of inheritance to deworming under which R_2 would have at least her due level of influence. Hence, there must be some region of the horizontal axis that exactly *coincides* with R_2 's shortfall function.

On the other hand, there must also be at least some less-than-100% allocations of inheritance to deworming under which R_2 would have less than her due influence. After all, we have stipulated that R_2 cares about money more than she cares about free time. Hence, the completion of PARTIAL NONCOMPLIANCE under which our decision maker is required to donate all of her inheritance to deworming must be *less* choiceworthy than the IDEAL PLAN according to R_2 . And from this it follows that there must be some less-than-100% allocations of inheritance to deworming under which R_2 's shortfall

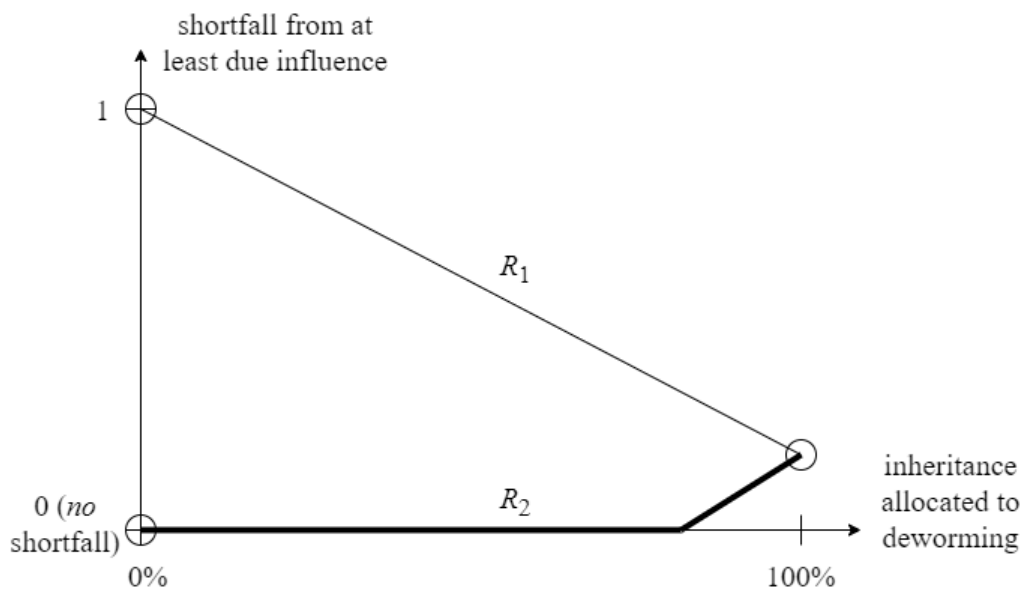


Figure 4.8: Both shortfall functions if each representative cares roughly the same about time and money

function must be greater than zero.

The exact shape of R_2 's shortfall function will (once again) depend on the extent to which R_2 cares about money more strongly than she cares about time. If R_2 cares only very slightly more about money, then R_2 can fall short of due influence only if a rather large amount of inheritance is allocated to deworming – as illustrated in figure 4.8. But, on the other hand, if R_2 cares much more about money, then R_2 can fall short of due influence even if only a modest amount of money is allocated to deworming – as illustrated in figure 4.9.

Regardless of exactly how much R_2 cares about money and time, we can always be certain that the nonzero segment of R_2 's shortfall function must

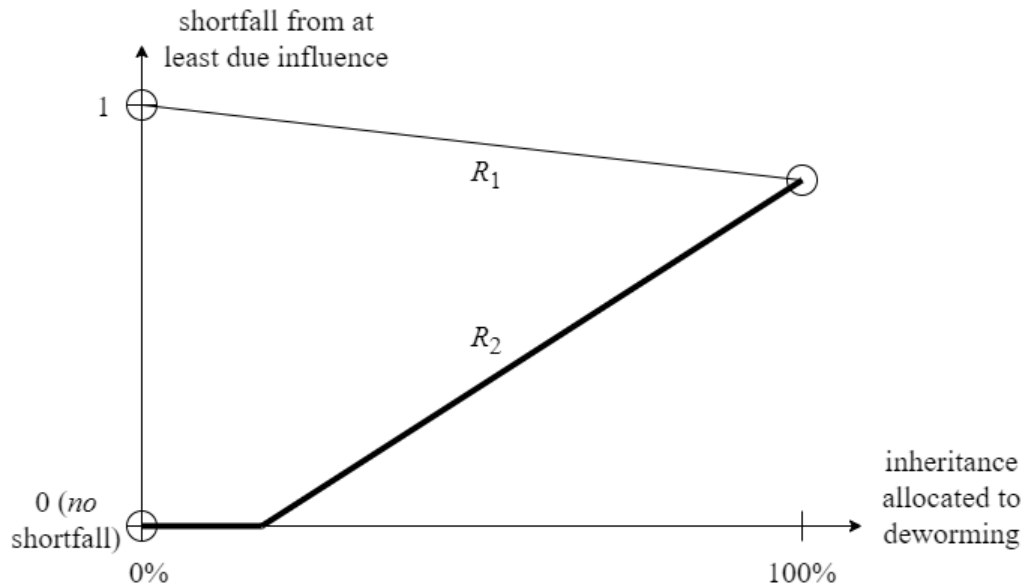


Figure 4.9: Both shortfall functions if each representative cares much more about one of our two resources

have a *steeper* gradient than that of R_1 's shortfall function. After all, I have stipulated that R_1 cares more about time than money, whereas by contrast R_2 cares more about money than time. Hence, varying our decision maker's allocation of her inheritance must alter R_1 's shortfall from due influence rather than strongly than it would alter R_2 's shortfall. Figures 4.8 and 4.9 both illustrate this difference.

Now that we know how R_1 and R_2 's shortfalls from due influence will depend upon how our decision maker allocates her inheritance, we can now ask which possible donation distribution would satisfy the SHORTFALL CRITERION. That is to say, we can now ask which possible donation distribution would *overall* come as close as possible to giving *all* of our theory represen-

tatives at least their due influence.

Our exact answer to this question will depend upon exactly how we decide to *aggregate* the individual shortfall measures for R_1 and R_2 into a single measure of the *overall* extent to which any given course of behaviour would fall short of giving these two representatives at least their due influence. I will discuss this choice point in §4.2.4 below.

That being said, even before we consider this choice point about aggregation, we can in fact conclude straightaway that the SHORTFALL CRITERION will not be satisfied by any possible completions of our decision maker's PARTIALLY NONCOMPLIANT course of behaviour under which R_2 would have greater than her due influence. After all, if our donation distribution gives R_2 greater than its due influence, then it is must be possible to shift at least a little more money over to deworming without giving R_2 any less than its due influence. And of course, shifting money over to deworming must decrease R_1 's shortfall from due influence. Hence, shifting a certain amount of money over to deworming must weakly dominate the original distribution with respect to the desideratum of giving each theory representative at least their due influence. Thus, the SHORTFALL CRITERION cannot be satisfied by any donation distribution that gives R_2 greater than due influence.

The proportion of possible distributions that we can now rule out on this basis will of course depend upon the extent to which R_2 cares about money much more strongly than she cares about time. To begin with, recall that if R_2 cares only very slightly more about money, then R_2 would have

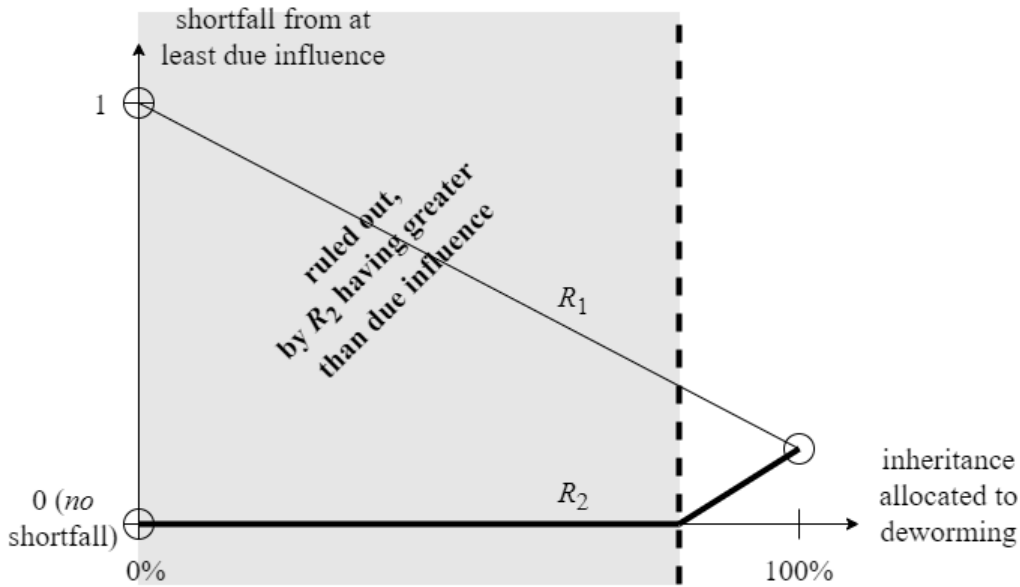


Figure 4.10: Distributions that give R_2 greater than her due influence, assuming that R_2 cares about money only slightly more than she cares about time

greater than her due influence under all possible donation distributions other than those which allocate close to 100% of our decision maker's inheritance to deworming. Thus, under these conditions we can be certain that any donation distribution will satisfy the SHORTFALL CRITERION for appropriate compensation only if this distribution requires our decision maker to donate most or all of her money to deworming. This result is illustrated in figure 4.10.

On the other hand, recall that if R_2 cares about money much more strongly than she cares about time, then R_2 would have greater than her due influence only under donation distributions which allocate at most a

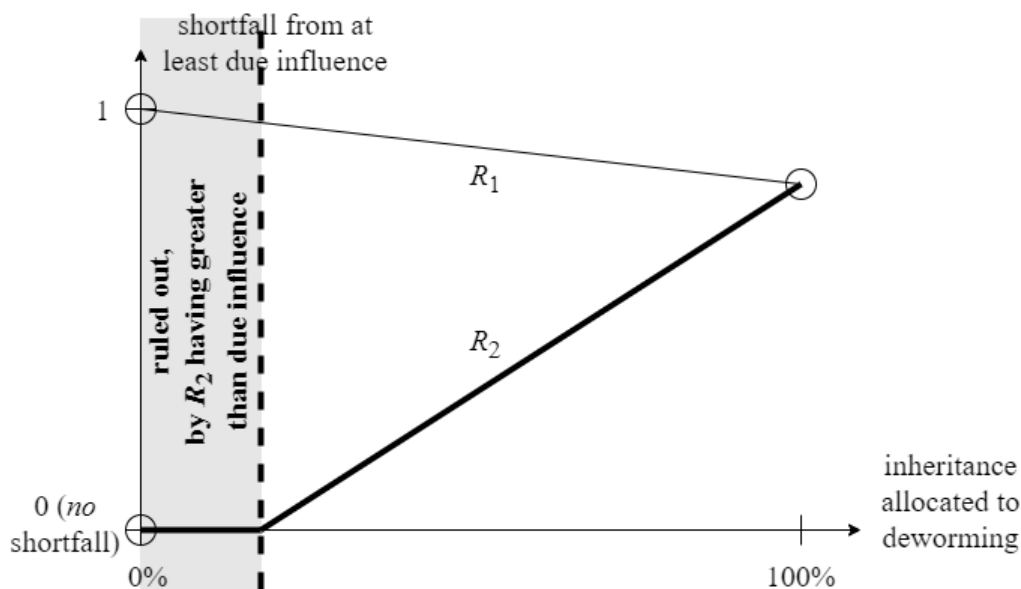


Figure 4.11: Distributions that give R_2 greater than her due influence, assuming that R_2 cares about money much more than she cares about time

small proportion of our decision maker's inheritance to deworming. Thus, under these conditions, any given donation distribution can be ruled out as inappropriate on the basis that it would give R_2 more than its due influence only if this distributions requires our decision maker to donate at most a small amount of money to deworming (as illustrated in figure 4.11). And in order to say anything more specific about the implications of my SHORTFALL CRITERION under these particular conditions, I now need to turn my attention to the question of how we should aggregate R_1 and R_2 's individual shortfall measures into a single measure of overall shortfall.

4.2.4 Overall shortfall

How should we aggregate our measurements of shortfall for each of our individual theory representatives into a measure of the overall extent to which these theory representatives fall short of indifference between the IDEAL and the PARTIALLY NONCOMPLIANT courses of behaviour? One possible option here would just be to *sum* over the individual shortfall measurements for all of our theory representatives – call this the *total-shortfall* aggregation proposal. Or, secondly, we could instead take the *maximum* over all individual shortfall measurements – call this the *greatest-shortfall* aggregation proposal. Or, thirdly, we could instead adopt any one of a whole number of more complicated alternative aggregation procedures.⁷

It is not obvious to me which of these shortfall aggregation proposals should be incorporated into COMPENSATIONISM. Fortunately, however, none of my claims about IMM in the remainder of this dissertation will depend upon this question. Hence, I will not attempt to conclusively settle this particular theoretical choice point here.

That being said, my own tentative preference would be for COMPENSATIONISTS to adopt the *total-shortfall* aggregation proposal. This is because the total-shortfall proposal allows COMPENSATIONISTS to avoid the risk of

⁷For instance, one potential compromise between the total- and greatest-shortfall aggregation proposals would be for us to take a *weighted sum* over the individual shortfall measurements for all of our theory representatives, where each representative's weight in this sum is proportional to her shortfall from indifference. We could describe this as a kind of 'prioritarian' aggregation procedure.

levelling-down theory representatives who have not yet been disadvantaged by past noncompliance.

I can illustrate this risk of levelling-down by once again using my running example of noncompliance in the **Inheritance** set of circumstances. More particularly, I will discuss what the implications of the total- and greatest-shortfall aggregation procedures would be under these conditions.

Consider, first of all, the greatest-shortfall aggregation proposal. Recall that – as illustrated in figures 4.8 and 4.9 above – R_1 's shortfall from at least due influence must always be at least as large as R_2 's shortfall, regardless of how our decision maker allocates her inheritance. Hence, the greatest shortfall will always just be R_1 's individual shortfall, and so our greatest shortfall function must always be minimised by our decision maker donating all of her inheritance to deworming, since this is the donation distribution which always minimises R_1 's individual shortfall. Thus, the greatest-shortfall version of COMPENSATIONISM would imply that the most appropriate option under these conditions of past noncompliance is for our decision maker to donate all of her inheritance to deworming.

By contrast, let us now consider the total-shortfall aggregation proposal. Recall that – as illustrated in figures 4.8 and 4.9 above – the nonzero segment of R_2 's shortfall function must always have a steeper gradient than that of R_1 's shortfall function, regardless of the exact extent to which R_2 cares about money more than time. Hence, the total shortfall function must always have a *positive* gradient whenever R_2 's individual shortfall is nonzero, as illustrated

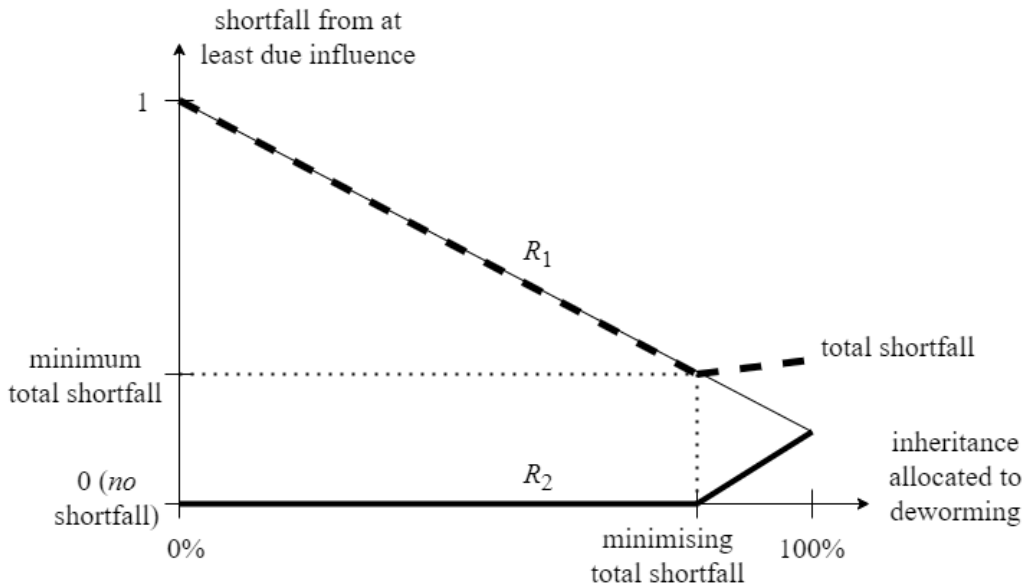


Figure 4.12: Total shortfall function if each representative cares roughly the same about time and money

in figures 4.12 and 4.13. Therefore, our greatest shortfall function must always be minimised by our decision maker allocating less than 100% of her inheritance to deworming, once again as illustrated in figures 4.12 and 4.13. Thus, the total-shortfall version of COMPENSATIONISM would imply that the most appropriate option under these conditions of past noncompliance is for our decision maker to donate less than 100% of her inheritance to deworming.

More specifically, total-shortfall COMPENSATIONISM would imply that the allocation which is most appropriate here should depend upon the extent to which each of our two theory representatives cares about one of the two resources more strongly than they care about the other. If each of these two theory representatives cared only very slightly more about one or the other of

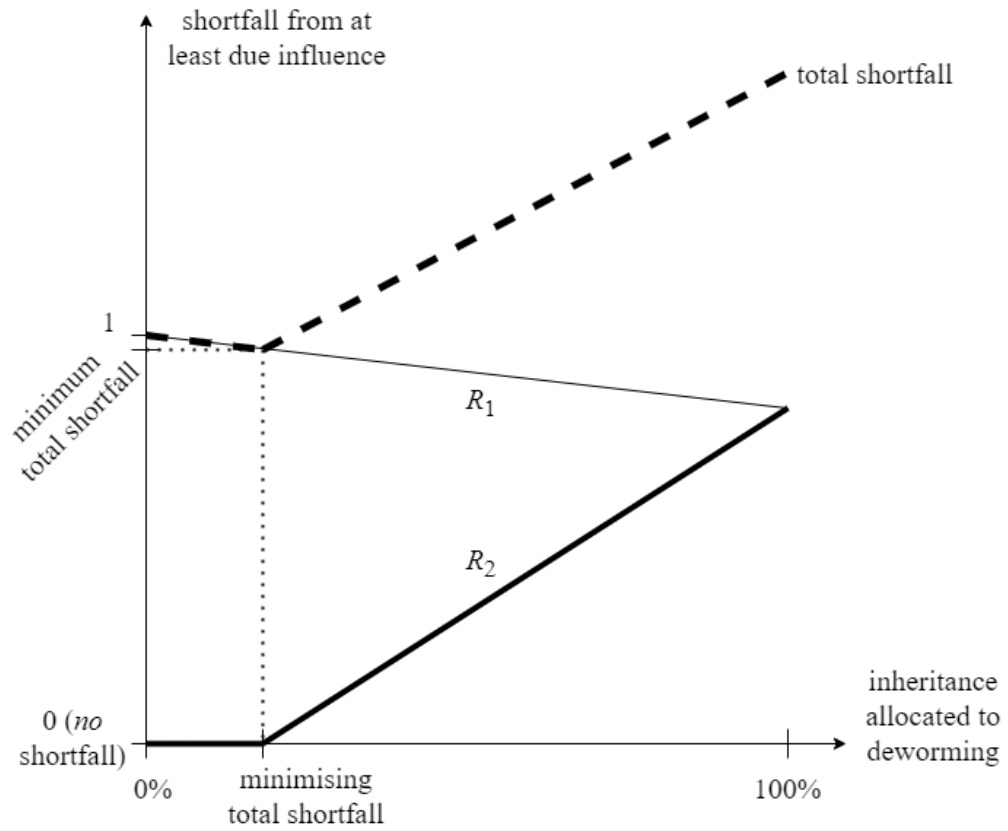


Figure 4.13: Total shortfall function if each representative cares much more about one of our two resources

these two resources, then total-shortfall COMPENSATIONISM would imply that it is most appropriate for our decision maker to donate only slightly less than 100% of her inheritance to deworming (as illustrated in figure 4.12). But, at the other extreme, if each of these two theory representatives care very much more about one or the other of these two resources, then total-shortfall COMPENSATIONISM would imply that it is most appropriate for our decision maker to donate only a small fraction of her inheritance to deworming (as illustrated in figure 4.13).

However, *regardless* of exactly which of these possibilities obtains, we can be certain that the appropriate amount of money to donate to deworming under these circumstances according to total-shortfall COMPENSATIONISM must be *less* than the appropriate amount according to the greatest-shortfall version of COMPENSATIONISM.⁸ In other words, total-shortfall COMPENSATIONISM recommends less-extensive compensation transfers than greatest-shortfall COMPENSATIONISM under these circumstances.⁹

This strikes me as being an important advantage of total- over greatest-shortfall COMPENSATIONISM. This advantage is especially vivid in versions of **Inheritance** wherein each of our two theory representatives cares very much more about one or the other of the two resources. Recall that, under these

⁸Recall that according to greatest-shortfall COMPENSATIONISM, under these circumstances it is most appropriate for our decision maker to donate all of her inheritance to deworming.

⁹That being said, there are other possible sets of circumstances in which the compensation transfers recommended by total-shortfall COMPENSATIONISM would be *more* extensive than those recommended by greatest-shortfall COMPENSATIONISM. But, for the sake of brevity, I will prescind from discussing these kinds of circumstances.

conditions, our decision maker donating all of her inheritance to deworming would cause both R_1 and R_2 to fall very far short of having at least their due influence.¹⁰ Hence, shifting from (1) this *greatest*-shortfall minimising allocation to (2) the *total*-shortfall minimising allocation would completely eliminate R_2 's individual shortfall – at the cost of only very slightly increasing R_1 's individual shortfall. Thus, in at least this example, greatest- but not total-shortfall COMPENSATIONISM is committed to an implausible kind of ‘levelling-down.’ This illustrates one important advantage of the total-shortfall aggregation proposal.

4.3 Descriptive updating

In addition to the question of how we should handle past noncompliance, another new question raised by the possibility of intertemporal bargaining concerns cases wherein our theory representatives initially all have highly inaccurate descriptive beliefs about the choice situations that their decision maker will confront in the future. For instance, imagine that in **Risky Inheritance** (first introduced in §4.1 above), our decision maker is initially almost certain that she will only inherit a *very small* amount of money from her grandmother. Hence, in IMM's model for this decision maker, the theory representatives R_1 and R_2 will both share this descriptive belief that she is very unlikely to inherit a large amount of money.

¹⁰After all, under these conditions R_1 cares very little about compensation in the form of money, whereas R_2 cares very greatly about losing control over money.

resource:	time	money
R_1 's preferred use:	nuclear disarmament	deworming
R_2 's preferred use:	orphanage	soup kitchens
representative that cares about this resource more than the other resource:	R_1	R_2
initial contract:	99.5% to disarmament 0.5% to orphanage	0% to deworming 100% to soup kitchens

Figure 4.14: The initial contracts in **Risky Inheritance**

Suppose that in light of these descriptive credences, R_1 and R_2 both agree to a contract under which R_2 will transfer to R_1 only *half* of R_2 's endowment of control rights over free time, in return for R_1 committing to transfer to R_2 *all* of R_1 's future endowment of control rights over however much money the decision maker eventually inherits. Hence, our decision maker will be instructed to split her free time 99.5:0.5 between disarmament campaigning and volunteering at the local orphanage, and then after that to donate all of her inheritance to soup kitchens. These instructions are summarized in figure 4.14.

However, now suppose that our decision maker in fact turns out to inherit a very large amount of money from her grandmother. Under these conditions, the terms of the contract between R_1 and R_2 would seem to imply that it is most appropriate for our decision maker to donate 100% of her large inheritance windfall to local soup kitchens, despite the fact that she only has 1% credence in the moral theory T_2 that favours donating to this charity.¹¹

¹¹Assume, for the rest of this section, that our decision maker's credence distribution

Unfortunately, this implication strikes many of us as quite implausible. It is intuitively inappropriate for a moral theory T_2 in which our decision maker only has 1% credence to have total control over her response to the highly consequential question of how to distribute her inheritance – for no better reason than the fact that R_2 ‘got lucky’ in having gambled on our decision maker receiving a large windfall.

One way to pump this intuition is to emphasise that in the IMM model for **Risky Inheritance**, R_1 might have agreed to transfer to R_2 all of R_1 ’s future endowment of control rights over the inheritance windfall only because our decision maker initially had very low-quality descriptive evidence about the amount of money that she would likely inherit. At the time when R_1 agreed to this transfer, the decision maker’s best evidence suggested that she would only inherit a very small amount of money. However, the decision maker can now recognise with hindsight that her earlier evidence was in this respect misleading (or at best incomplete), because she has now inherited a very large amount of money. Hence, any naive version of IMM which says that it is most appropriate for our decision maker to donate everything to soup kitchens is in effect claiming that this decision maker’s use of her inheritance should be determined by an IMM model incorporating only the low-quality profile of descriptive evidence that the decision maker happened to have at some earlier point in time before she received her inheritance. However, it

over moral theories never changes over the course of her lifetime. I will relax this assumption in §4.4 below.

seems to many of us that the decision maker's use of her inheritance should not be determined in this way by the 'dead hand' of her past ignorance. Let's call this the *problem of descriptive updating*.

Fortunately, a more sophisticated version of IMM can avoid any intuitively implausible implications in all cases of descriptive updating. Roughly stated, the basic idea behind this more sophisticated version of IMM is that at any given moment in time t , our decision maker should ask herself which contract(s) her theory representatives *would* have earlier agreed to if only this decision maker had always had the descriptive credence distribution which she actually has at time t . Let us call this the t -DESCRIPTIVE CREDENCES profile of counterfactual contracts. Then according to my preferred version of IMM, this t -DESCRIPTIVE CREDENCES profile of contracts should in some sense be used to determine how it is most appropriate for our decision maker to behave at time t . Note that this last sentence is intended to be somewhat vague, since I plan to discuss two possible precisifications of this basic idea later in the present section.

Before that, however, let me first of all discuss how to apply this basic idea to our **Risky Inheritance** descriptive updating example. Roughly stated, my preferred version of IMM implies that once our decision maker has learned the true amount of money that she will inherit from her grandmother, then at this point in time our decision maker should ask herself which contract(s) R_1 and R_2 would have earlier agreed to if only these two theory representatives had known from the beginning what this inheritance amount would be.

Let us call this the HIGH-INFORMATION profile of counterfactual contracts. Then according to my preferred version of IMM, this HIGH-INFORMATION profile of contracts should in some sense be used to determine how it is most appropriate for this decision maker to use her inheritance.¹²

Which (if any) contracts would R_1 and R_2 have agreed to if they had known all along that our decision maker would inherit a very large amount of money? Well, if R_1 and R_2 had known from the beginning what this inheritance amount would be, then it is safe to assume that R_1 would never have agreed to transfer to R_2 *all* of R_1 's future endowment of control rights over inheritance. In particular, let us suppose that if R_1 and R_2 had known the inheritance amount all along, then R_2 would have agreed to transfer to R_1 *all* of R_2 's endowment of control rights over free time, in return for R_1 agreeing to transfer to R_2 control rights over only 1% of the decision maker's inheritance – leaving R_2 with control over $1\% + 1\% = 2\%$ of that inheritance windfall. Hence, under this HIGH INFORMATION contract, our decision maker would have been instructed to spend all of her free time campaigning for nuclear disarmament, and then after that to split her inheritance 98:2 between deworming and soup kitchens. These instructions are summarized in figure 4.15.

According to my preferred version of IMM, these HIGH INFORMATION instructions should in some sense determine how it is most appropriate for

¹²Once again, remember that this last sentence is intended to be somewhat vague, since I plan to discuss two possible precisifications below.

resource:	time	money
R_1 's preferred use:	nuclear disarmament	deworming
R_2 's preferred use:	orphanage	soup kitchens
representative that cares about this resource more than the other resource:	R_1	R_2
contract under ignorance:	99.5% to disarmament 0.5% to orphanage	0% to deworming 100% to soup kitchens
HIGH-INFORMATION contract:	100% to disarmament 0% to orphanage	98% to deworming 2% to soup kitchens

Figure 4.15: Two contracts in **Risky Inheritance**

our decision maker to use her inheritance. However, let me state clearly here that these HIGH-INFORMATION instructions should *not* in any sense be taken to determine how it is or was appropriate for our decision maker to have spent her earlier free time. On the contrary, how it is or was appropriate for our decision maker to spend her free time should be determined by the profile of contracts that R_1 and R_2 would have agreed to when they believed that our decision maker would only inherit a very small amount of money from her grandmother. Thus, according to my preferred version of IMM, it is *timelessly* true that the most appropriate use for our decision maker's free time is or was a 99.5:0.5 split between disarmament campaigning and orphanage volunteering.¹³ Likewise, it is also timelessly true that most appropriate use for our decision maker's inheritance should in some sense be determined by the HIGH-INFORMATION counterfactual.

¹³More generally: how it was appropriate for our decision maker to have behaved at some earlier point in time t must in some sense be determined only by this decision maker's credences at that earlier point in time t .

This concludes my discussion of how it would have been most appropriate for our decision maker to have distributed her free time at the start of **Risky Inheritance**. Having concluded this discussion, I will now turn my attention to the question of how it would now be most appropriate for our decision maker to distribute her inheritance. Recall that according to my preferred version of IMM, this appropriateness verdict should in some sense be determined by the HIGH-INFORMATION counterfactual instructions. However, as I have already remarked, my framing of this idea has thus far been deliberately vague. Hence, I shall now introduce two potential precisifications of this vaguely-stated basic idea.

(1) NONINTERACTIONISM: according to the NONINTERACTIONIST precisification, the most appropriate donation distribution under these conditions must simply be the 98:2 split between deworming and soup kitchens which our decision maker would have been instructed to realize in the HIGH-INFORMATION counterfactual. Crucially, NONINTERACTIONISM implies that this 98:2 split is the most appropriate inheritance distribution under these conditions *regardless* of how our decision maker previously used her free time. Thus, NONINTERACTIONISM is to some extent analogous to the NONCONTAMINATIONIST response to noncompliance (introduced in §4.2.1 above).

(2) HOLISM: by contrast, according to the HOLIST precisification, a 98:2 split between deworming and soup kitchens would

be the most appropriate donation distribution under these conditions *only if* our decision maker earlier spent all of her free time campaigning for nuclear disarmament, as she would have been instructed to in our HIGH-INFORMATION counterfactual. More generally, HOLISM implies that it is most appropriate for our decision maker to select the inheritance distribution ϕ which would minimise her shortfall from rendering every theory representative at worst indifferent between (a) the total lifetime course of action which our decision maker would have been instructed to select under the HIGH-INFORMATION counterfactual, and (b) what our decision maker's total lifetime course of action will in fact be, supposing only that she distributes her inheritance in accordance with ϕ .¹⁴ Thus, HOLISM is to some extent analogous to the COMPENSATIONIST response to noncompliance (introduced in §4.2.1 above).¹⁵

I have suggested that NONINTERACTIONISM and HOLISM are to some extent analogous to the NONCONTAMINATIONIST and COMPENSATIONIST responses to noncompliance. However, let me also state clearly here that NONINTERACTIONISM and HOLISM should *not* be thought of as being mere appli-

¹⁴Recall that I discussed the notion of shortfall from at worst indifference in §§4.2.3-4.2.4 above.

¹⁵A third possible precisification of my preferred response to descriptive updating could be analogous to the NULLIFICATIONIST response to noncompliance. However, this precisification would suffer from some fatal problems which would be closely analogous to the fatal problems suffered by NULLIFICATIONISM (recall §4.2.1 above).

cations of special cases of NONCONTAMINATIONISM and COMPENSATIONISM. COMPENSATIONISM, for example, is a proposal for handling cases of past noncompliance which says that it can be most appropriate for our decision maker to choose options not recommended by her IDEAL PLAN in cases where some of her earlier choices were *inappropriate*. By contrast, HOLISM is a proposal for handling cases of descriptive updating, which says that it can be most appropriate for our decision maker to choose options not recommended by her HIGH-INFORMATION profile of instructions in cases where some of her earlier choices would themselves not have been recommended by the HIGH-INFORMATION profile of instructions. Hence, HOLISM implies that it can sometimes be most appropriate for our decision maker to deviate from her HIGH-INFORMATION instructions even in cases where all of her earlier choices were *appropriate*. After all, recall that according to my preferred response to descriptive updating, the HIGH-INFORMATION profile of instructions should in some sense determine how it is most appropriate for our decision maker to behave *only after* this decision maker has in fact updated her descriptive credences. Hence, our decision maker having previously chosen some options that would not have been recommended by her HIGH-INFORMATION profile of counterfactual instructions can be totally compatible with this decision maker having always acted appropriately.

In fact, the **Risky Inheritance** set of circumstances is one example of this phenomenon. Recall that in **Risky Inheritance**, it is a timeless truth that it would be most appropriate for our decision maker to split her

free time 99.5:0.5 between disarmament campaigning and orphanage volunteering. But, in the HIGH-INFORMATION counterfactual world, this decision maker would by contrast have been instructed to spend all of her free time campaigning for disarmament. Hence, our decision maker having previously chosen a division of her free time which would not have been recommended by her HIGH-INFORMATION profile of counterfactual instructions can be totally compatible with this decision maker having used her free time appropriately.

Imagine that our decision maker did in fact distribute her free time appropriately, splitting it 99.5:0.5 between disarmament campaigning and orphanage volunteering. Then under these conditions, HOLISM implies that it is now most appropriate for this decision maker to donate slightly *more* than 98% of her inheritance to deworming, with the remaining *less* than 2% being donated to soup kitchens. (For the sake of brevity, I will prescind from explaining how HOLISM implies this result.)

Exactly which inheritance distribution would be most appropriate under these conditions according to HOLISM will of course depend upon exactly how we measure overall shortfalls from at worst due influence. And as I discussed in §§4.2.3-4.2.4 above, there are several complicated choice points associated with measuring overall shortfalls.

However, regardless of exactly how we settle those complicated choice points, under these conditions we can be certain that the HOLIST version of my preferred response to descriptive updating will recommend for our decision maker to donate more than 98% of her inheritance to deworming (since

resource:	time		money	
	disarmament	orphanage	deworming	soup kitchens
contract under ignorance:	99.5%	0.5%	0%	100%
decision maker's actual course of action:	99.5%	0.5%	?	?
appropriate according to NONINTERACTIONISM:	99.5%	0.5%	98%	2%
appropriate according to HOLISM:	99.5%	0.5%	> 98%	< 2%

Figure 4.16: Courses of behaviour in **Risky Inheritance**

merely giving 98% would be to make no adjustments at all, contrary to what HOLISM recommends). Thus, under these conditions we can be certain that HOLISM will disagree with the NONINTERACTIONIST recommendation that it would be most appropriate for our decision maker to split her inheritance 98:2 between deworming and soup kitchens. These two different recommendations are summarized in figure 4.16.

NONINTERACTIONISM and HOLISM hence issue subtly different recommendations under these conditions. Fortunately, however, each of these views avoids the ‘problem of descriptive updating’ introduced at the beginning of this section. Recall that the challenge to IMM posed by my running example of descriptive updating in **Risky Inheritance** involved avoiding the implausible result that it would be most appropriate for our decision maker to donate all of her inheritance to soup kitchens, even after she has learned that her inheritance is far larger than she originally expected. And NONINTERACTIONISM and HOLISM both quite clearly avoid this implausible result, since they agree that it would be inappropriate for our decision maker to

donate any more than 2% of her inheritance to soup kitchens. This strikes me as an attractive response to these particular circumstances of descriptive updating.

For the remainder of this section, I will now turn my attention to the question of how we should not choose between NONINTERACTIONISM and HOLISM. I will not attempt to conclusively settle this particular choice point here. However, I will present at least some considerations in favour of HOLISM.

I have already pointed out that NONINTERACTIONISM and HOLISM should be conceived of being mere applications of the NONCONTAMINATIONIST and COMPENSATIONIST responses to noncompliance. That fact notwithstanding, however, the motivating ideas behind NONINTERACTIONISM and HOLISM are natural extensions of the motivating ideas behind NONCONTAMINATIONISM and COMPENSATIONISM, first introduced in §4.2.1 above. HOLISM suggests that the appropriateness of any particular option should depend upon the extent to which choosing that option would bring our decision maker's total lifetime course of action closer to giving each moral theory its due influence *as defined relative to the decision maker's present descriptive credences*. Furthermore, NONINTERACTIONISM suggests that the appropriateness of any particular option should depend upon whether this option is part of the total lifetime HIGH-INFORMATION plan of action which would have come as close as possible to giving each moral theory its due influence – once again *as defined relative to the decision maker's present descriptive credences*.

Hence, the debate between HOLISM and NONINTERACTIONISM is very

closely analogous to the debate between COMPENSATIONISM and NONCONTAMINATIONISM (recall §4.2.1 above).¹⁶ On the one hand, HOLISM says that an option's appropriateness should *directly* depend upon the desideratum of due influence (as defined relative to present descriptive credences). And, on the other hand, NONINTERACTIONISM says that an option's appropriateness should directly depend upon the HIGH-INFORMATION lifetime plan of action, and hence should depend only *indirectly* upon the desideratum of due influence (once again, as defined relative to present descriptive credences).

So understood, HOLISM strikes me as a more attractive view than NONINTERACTIONISM. If overall due influence is really a legitimate desideratum, then it seems to me most plausible to suppose that the appropriateness of any given option should *directly* depend upon how these options stand with respect to the due influence desideratum. At the very least, NONINTERACTIONISTS owe us an explanation of why an option's appropriateness should depend only indirectly upon the desideratum of due influence.

Hence, my preferred version of IMM combines a HOLIST response to descriptive updating with the COMPENSATIONIST response to noncompliance. According to this version of IMM:

the most appropriate course of action for our decision maker to

¹⁶As a matter of fact, the difference between HOLISM and NONINTERACTIONISM can actually be defined in terms of the difference between COMPENSATIONISM and NONCONTAMINATIONISM. The most appropriate option according to HOLISM (or, respectively: NONINTERVENTIONISM) is always the option which *would* have been appropriate according to COMPENSATIONISM (or, respectively: NONCONTAMINATIONISM) if only our decision maker had always had her current descriptive credences.

choose at any given moment in time t is the course of action ϕ which would minimise her shortfall from rendering every theory representative at worst indifferent between (a) the total lifetime course of action that our decision maker would have been instructed to select if she had always had the descriptive credence distribution which she has at time t , and (b) what the decision maker's total lifetime course of behaviour will in fact be, supposing only that she will follow ϕ in all future choice situations.

To summarize, I have argued in this section that whenever our decision maker updates her descriptive credences, she should ask herself which trades and contracts her theory representatives would have agreed to had this decision maker always known everything that she knows now. This approach to handling cases of descriptive updating maintains all of the core features and advantages of the IMM response to moral uncertainty, but at the same time also ensures that no low credence moral theories can ever have an inappropriately outsized level of influence over our decision maker's total lifetime course of behaviour. Thus, this strikes me as a highly attractive response to the problem of descriptive updating.

4.4 Moral updating

Having now discussed how IMM should handle descriptive updating, I will next turn my attention to the question of how IMM should handle cases

wherein our decision maker updates her *moral* credence distribution.

For instance, imagine that in the IMM model for **Inheritance**, R_1 and R_2 both agree to a contract under which R_2 will transfer to R_1 all of R_2 's initial endowment of control rights over the decision maker's free time, in return for R_1 agreeing to transfer to R_2 all of R_1 's future endowment of control rights over the decision maker's inheritance. Under this assumption, IMM would imply that it is most appropriate for our decision maker to spend all of her free time campaigning for nuclear disarmament (as favoured by R_1). Hence, let us imagine that our decision maker does in fact spend all of her free time campaigning for disarmament.

However, let us now also imagine that before our decision maker receives her inheritance, she happens to update her credence distribution over moral theories. In particular, imagine that her credence in T_1 drops from 50 to 0%, with her credence in a third moral theory T_3 correspondingly rising from 0 to 50%.¹⁷

In the IMM bargaining model for this decision maker, it is natural to suppose that T_2 and T_3 's representatives R_2 and R_3 should eventually each be endowed with control over 50% of our decision maker's inheritance, and hence that T_1 's representative R_1 should not be endowed with control over any of the inheritance. After all, at the moment in time when our decision

¹⁷Updating from 0 to 50% is of course impossible under orthodox Bayesian conditionalization. However, it wouldn't make any difference to my arguments in the rest of this section if I instead assumed that our decision maker initially has some low but nonzero credence in the moral theory T_3 . Hence, for simplicity of exposition, it will do no harm for me to simply assume that this credence updates from 0 to 50%.

maker actually receives her inheritance, I have stipulated that she will have zero credence whatsoever in the moral theory T_1 . Hence, splitting control over the inheritance in proportion to the decision maker's moral credences at that moment in time would clearly be equivalent to splitting these control rights 50:50 between R_2 and R_3 .

However, if R_1 is not endowed with any control rights over this decision maker's inheritance, then it will be impossible for R_1 to honour her contractual commitment to transfer control rights over 50% of this inheritance to R_2 . Hence, under these conditions R_2 will not be entitled to any additional control rights over our decision maker's inheritance beyond her 50% initial endowment. Thus, the IMM approach to moral uncertainty would seem to imply that it is now most appropriate for our decision maker to donate 50% of her inheritance to each of T_2 and T_3 's favoured uses for it.

The intuitive plausibility of this possible recommendation will depend upon exactly how we specify the details of our new moral theory T_3 . On the one hand, imagine first of all that T_3 (i) shares T_1 's view that our decision maker should donate as much of her inheritance as possible to the deworming charity, but at the same time (ii) shares T_2 's view that the decision maker should spend as much of her free time as possible volunteering at the local orphanage (as summarized in figure 4.17). In other words, let us assume that R_2 and R_3 agree with each other that our decision maker should not have spent her free time campaigning for nuclear disarmament, but by contrast disagree with each other about how our decision maker should distribute her

	resource:	time	money
R_1	favours: controls:	nuclear disarmament 50%	deworming 0%
R_2	favours: controls:	orphanage 50%	soup kitchens 50%
R_3	favours: controls:	orphanage 0%	deworming 50%
decision maker's actual use:		nuclear disarmament	?

Figure 4.17: Choices and preferences in **Inheritance**

inheritance.

Under these conditions, it *would* strike me as plausible to suppose that it is now most appropriate for our decision maker to split her inheritance 50:50 between soup kitchens and deworming. After all, this is now the only available course of action that would give T_2 and T_3 equal influence over our decision maker's actual lifetime course of behaviour.

Unfortunately, however, there are some other possible specifications of T_3 under which this 50:50 split between soup kitchens and deworming would not strike me as appropriate. For instance, let us now imagine that T_3 is extremely similar to T_1 . In particular, suppose that T_3 agrees with T_1 (and hence disagrees with T_2) about how our decision maker should use her free time and her inheritance. In other words, let us assume that R_3 agrees with R_1 that our decision maker was right to have spent her free time campaigning for nuclear disarmament, and also that she should now donate as much of her inheritance as possible to deworming.

Under these new conditions, it would strike me as inappropriate for our decision maker to split her inheritance 50:50 between soup kitchens and de-

worming. If our decision maker still had 50% credence in each of the two moral theories T_1 and T_2 , then IMM implies that it would still be most appropriate for this decision maker to donate all of her inheritance to soup kitchens. But, *ex hypothesi*, T_3 is extremely similar to T_1 , and totally agrees with T_1 about how our decision maker should behave in **Inheritance**. Hence, it is implausible to suppose that our decision maker shifting 50% of her credence mass from T_1 to T_3 could shift the most appropriate inheritance distribution between deworming and soup kitchens from 0:100 to 50:50. Let's call this the *problem of moral updating*.

Fortunately, a more sophisticated version of IMM can avoid any intuitively implausible implications in cases of moral updating. In fact, we can avoid these implausible implications simply by handling moral updating strictly analogously to how I suggested that we should handle descriptive updating in §4.3 above. Roughly stated, the basic idea here is that at any given moment in time t , our decision maker should ask herself which contract(s) her theory representatives *would* have earlier agreed to if only this decision maker had always had the moral distribution which she actually has at time t . Let us call this the t -MORAL CREDENCES profile of counterfactual contracts. Then according to my preferred version of IMM, this t -MORAL CREDENCES profile of contracts should in some sense be used to determine how it is most appropriate for our decision maker to behave at time t . Once again (recall §4.3 above), note that this last sentence is intended to be somewhat vague, since it is designed to be compatible with both a 'NONINTERACTIONIST' and

a ‘HOLIST’ precisification.¹⁸

In the particular example of **Inheritance**, my preferred version of IMM says that when our decision maker changes her moral credence distribution, she should then ask herself what (if any) contract R_2 and R_3 would have earlier agreed to had the decision maker always had 50% credence in each of T_2 and T_3 . But of course, our answer to this question will depend upon how exactly we specify the details of our new moral theory T_3 .

First of all, consider the case in which T_3 is extremely similar to T_1 , and hence totally agrees with T_1 about how our decision maker should behave in **Inheritance**. Under these conditions, we can safely assume that R_2 would have agreed to a contract with R_3 exactly identical to her contract with R_1 . In other words, R_2 and R_3 would have agreed to a contract under which R_2 would have transferred to R_3 all of R_2 's initial endowment of control rights over the decision maker's free time in **Inheritance**, in return for R_3 agreeing to transfer to R_2 all of R_3 's future endowment of control rights over the decision maker's inheritance windfall. Hence, under these conditions, our theory representatives would have jointly instructed this decision maker to spend all of her free time campaigning for nuclear disarmament, and then to donate all of her inheritance to soup kitchens.

My preferred version of IMM implies that this counterfactual profile of instructions should in some sense determine which inheritance distribution is

¹⁸These two possible precisifications will obviously be very closely analogous to the ‘NONINTERACTIONIST’ and ‘HOLIST’ responses to descriptive updating introduced in §4.3 above.

now most appropriate. Moreover, I have stipulated that our decision maker did in fact spend all of her free time campaigning for nuclear disarmament, as required by this counterfactual profile of instructions. Hence, under these conditions, any reasonable precisification of my preferred version of IMM (be it NONINTERACTIONIST *or* HOLIST) will imply that it would be most appropriate for our decision maker to allocate all of her inheritance to soup kitchens (as favoured by R_2). This strikes me as a plausible response to these circumstances of moral updating.

My preferred version of IMM also has plausible implications in the case where we imagine that T_3 agrees with T_1 in favouring deworming, but also agrees with T_2 in favouring orphanage volunteering (as illustrated in figure 4.17 above). Under these conditions, R_2 and R_3 completely agree with each other about how the decision maker should distribute her free time in **Inheritance**, but by contrast have diametrically opposed preferences over how she should distribute her inheritance windfall. Hence, there are no opportunities for gains from trade or contract between R_2 and R_3 under these conditions. Thus, if our decision maker had always had 50% credence in each of the two moral theories T_2 and T_3 , then the representatives R_2 and R_3 would have initially jointly instructed this decision maker to spend all of her free time volunteering at the local orphanage. And then once this decision maker received her inheritance, R_2 would have instructed her to donate 50% to soup kitchens, whereas R_3 would have instructed her to donate the other 50% to deworming.

Once again, my preferred version of IMM implies that the counterfactual profile of instructions should in some sense determine which inheritance distribution is now most appropriate. Moreover, let us assume for the sake of simplicity that R_2 and R_3 's bargaining positions in **Inheritance** would be structurally identical, in the sense defined in §2.4 above. Then, under these conditions, every reasonable precisification of my preferred version of IMM (be it NONINTERACTIONIST *or* HOLIST) will imply that it would be most appropriate for our decision maker to split her inheritance 50:50 between de-worming and soup kitchens, *regardless* of how this decision maker spent her free time earlier. Once again, this strikes me as a plausible response to these circumstances of moral updating.

In all of the version of **Inheritance** which I have considered thus far in this section, the implications of my preferred approach to moral updating have never depended upon exactly how we precisify the vaguely-stated basic idea that how it is most appropriate for our decision maker to behave at any given moment in time t should in some sense be determined by her t -MORAL CREDENCES profile of counterfactual contracts. However, there are many other moral updating sets of circumstances in which the implications of my preferred approach to moral updating *would* depend upon exactly how we precisify this vaguely-stated basic idea. Hence, IMM faces yet another theoretical choice point.

In response to this particular theoretical choice point, my own view is that we should adopt a HOLIST response to moral updating. However, I will

not rehearse my arguments for this view here, since they would be almost identical to the arguments from §4.3 above in favour of the HOLIST response to descriptive updating.

Overall, then, my preferred version of IMM combines a HOLIST response to moral and descriptive updating with the COMPENSATIONIST response to noncompliance. Hence, according to my preferred version of IMM:

the most appropriate course of action for our decision maker to choose at any given moment in time t is the course of action ϕ which would minimise her shortfall from rendering every theory representative at worst indifferent between (a) the total lifetime course of action that our decision maker would have been instructed to select if she had always had the moral and descriptives credences which she has at time t , and (b) what the decision maker's total lifetime course of behaviour will in fact be, supposing only that she will follow ϕ in all future choice situations.

To summarize, I have argued in this section that whenever our decision maker updates her moral credences, she should ask herself which trades and contracts her theory representatives would have agreed to had our decision maker always had her current credence distribution. This approach to handling cases of moral updating maintains all of the core features and advantages of IMM, and also elegantly coheres with the approach to handling cases of descriptive updating that I defended in §4.3 above. Thus, this approach strikes me as a highly attractive response to the problem of moral updating.

Chapter 5

Discrete choice

I have demonstrated in §2 above how IMM can be applied in choice situations where the decision maker has to decide how she should distribute resources like time or money, given that the control rights over these kinds of resources can be divided among the theory representatives in proportion to credences. However, I have not yet discussed choice situations that concern something other than the distribution of resources. But this is an important lacuna, since many choice situations fall into this category. Just to take a stylized example, consider the following choice situation:

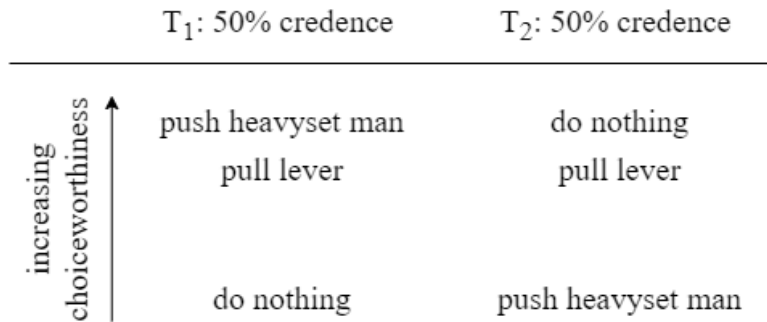
Trolley: a runaway trolley is headed towards five people who are tied to the tracks. The decision maker in this scenario has three options:

(i) pushing a heavysset man into the path of the trolley, intentionally killing him but saving the five;

- (ii) pulling a lever to redirect the trolley onto a side track where only two people are trapped, foreseeably killing them but saving the five;
- (iii) doing nothing, allowing the five to die.

Furthermore, suppose that this decision maker has 50% credence in a consequentialist moral theory T_1 , and 50% credence in a deontological moral theory T_2 . According to T_1 , whenever our decision maker confronts a choice situation like **Trolley**, it will be most choiceworthy to push the heavysset man, almost as choiceworthy to pull the lever, and highly unchoiceworthy to do nothing. Conversely, according to T_2 , whenever our decision maker confronts a choice situation like **Trolley**, it will be most choiceworthy to do nothing, almost as choiceworthy to pull the lever, and highly unchoiceworthy to push the heavysset man (illustrated in figure 5.1).

The decision maker in this **Trolley** scenario faces a choice between only three possible options (pushing the heavysset man, pulling the lever, and doing nothing). Thus, IMM cannot model **Trolley** as a choice situation in which the decision maker has to choose how to distribute some resource that can be divided 50-50 between the two theory representatives R_1 and R_2 . For instance, it obviously would not make any sense to model ‘the heavysset man’ or ‘the lever’ as resources that could be initially divided 50-50 between R_1

Figure 5.1: Choiceworthiness schedules in **Trolley**

and R_2 in the **Trolley** scenario. Instead, we need to extend IMM to show how it can model these kinds of ‘*discrete choice*’ situations. Fortunately, I will argue in this chapter that there exists at least one attractive response to this new challenge for IMM.

5.1 Divisible sequences

There are certain special cases in which it is not so difficult to see how we could specify an attractive IMM response to discrete choice scenarios like **Trolley**. In particular, let us temporarily suppose that the decision maker knows she is in fact about to confront a sequence of ten identical instances of the **Trolley** choice situation. In other words, after the decision maker pushes the heavyset man, pulls the lever, or allows the five to die, she will immediately find herself confronting another instance of the **Trolley** choice situation, with another trolley headed towards another five people tied to another set of tracks, another heavyset man standing nearby, and another

two people tied to a side track onto which the trolley can be redirected. Furthermore, immediately after the decision maker resolves this second **Trolley** situation, she will then face a *third* such choice situation; and so and so forth until the sequence of **Trolley** choice situations confronting her is over. And, as we always do, let us again assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations other than this sequence of ten in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contracts involving other scenarios in addition to **Trolley**.

Let us also suppose that according to both T_1 and T_2 , the choiceworthinesses of the three options in each of these **Trolley** situations do not depend on the situation's position in the overall sequence, or upon how the decision maker has behaved in any **Trolley** situations she has already encountered. Under this assumption, the discrete control right to determine how the decision maker will behave in any given instance of **Trolley** within the overall sequence of **Trolley** choice situations that she know she is about to confront can be treated like one indivisible unit of a larger overall resource with constant returns to scale (*viz.* control over how the decision maker will behave in the entire sequence of **Trolley** choice situations). In particular, if we suppose that the decision maker knows she is about to confront an *even* number of identical instances of the **Trolley** choice situation, then control over this sequence of choice situations could be divided without remainder 50-50 between R_1 and R_2 (in proportion to the decision maker's credences

in T_1 and T_2). For example, if the decision maker knows that she is about to confront ten identical instances of **Trolley**, then our initial assignment of control rights could endow R_1 with the right to choose how the decision maker will behave in the first five of these choice situations, but endow R_2 with the right to choose how the decision maker will behave in the remaining five choice situations.

After this initial endowment of control rights, R_1 and R_2 would then have the option to trade or make contracts with each other. Are any mutually beneficial trades or contracts available to the two representatives under these circumstances?

Imagine, first of all, an outcome in which R_1 and R_2 do not make any trades or contracts with each other. Under these conditions, it will be in R_1 's best interests to instruct the decision maker to push the heavysset man in each of the initial five **Trolley** situations; and it will be in R_2 's best interests to instruct the decision maker to do nothing in the remaining five **Trolley** situations. Unfortunately, neither R_1 nor R_2 is likely to be particularly happy with this sequence of choices, since each representative thinks that it would require the decision maker to choose a highly unchoiceworthy option in five of the total ten **Trolley** choice situations.

Now imagine, however, an outcome in which R_1 and R_2 both agree to instruct the decision maker to pull the lever in every instance of **Trolley**. Since R_1 and R_2 both think that pulling the lever is almost as choiceworthy as their favoured option in each **Trolley** situation, it is safe to assume that

both theory representatives will regard this overall sequence of choices as a significant improvement over the sequence of choices in which the decision maker pushes the heavysset man in five out of ten cases, but then does nothing in the remaining five cases. In other words, it makes sense to assume that pushing five times and then doing nothing five times is Pareto dominated by (among other things) pulling the lever in all ten cases. Since no rational economic agents with the ability to make contracts would ever settle for a Pareto-dominated outcome, IMM suggests that the decision maker pushing five times and then doing nothing five times is inappropriate in a sequence of ten identical instances of the **Trolley** choice situation.

This discussion raises the question of which overall plan of action would be *most* appropriate in our sequence of ten **Trolley** situations. At the moment – to again repeat a familiar refrain – we are missing several pieces of information that we would need in order to answer this question definitively. I have not yet specified in any detail how IMM should model the inter-representative bargaining process; and I have not yet specified exactly how T_1 and T_2 compare the choiceworthinesses of the three options available in **Trolley**. Since I will defer discussion of inter-representative bargaining until §7 of this dissertation, my definitive response to the sequence of ten **Trolley** situations will have to wait until then.

This point notwithstanding, there are nonetheless some ways of stipulating T_1 and T_2 's views about choiceworthiness in **Trolley** under which *any* plausible model of inter-representative IMM bargaining will imply that the

most appropriate outcome in our sequence of ten **Trolley** situations is for the decision maker to pull the lever in all ten choice situations. Once again (recall §§2.4-2.5 above), I have in mind cases in which R_1 's bargaining position vis-à-vis R_2 is structurally identical to R_2 's bargaining position vis-à-vis R_1 .

For example, suppose that according to T_1 , pushing the heavysset man has a choiceworthiness value of $+1 T_1$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_1$ -units, and doing nothing has a negative choiceworthiness value of $-1 T_1$ -units. Symmetrically, suppose that according to T_2 , doing nothing has a choiceworthiness value of $+1 T_2$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_2$ -units, and pushing the heavysset man has a negative choiceworthiness value of $-1 T_2$ -units.

Under these assumptions, R_1 and R_2 's bargaining positions are structurally identical to each other in the sense defined in §2.4 above. In fact, R_1 and R_2 's bargaining positions in this case closely parallel R_1 and R_2 's bargaining positions in the identical-positions precisification of **Three Charities** that I discussed in detail in §2.4. In §2.4, I demonstrated that in identical-positions precisifications of **Three Charities**, any plausible version of IMM will imply that it is most appropriate for the decision maker to donate her entire fortune to the compromise charity that is second best according to both of the moral theories in which the decision maker has positive credence. Thus, for reasons that strictly parallel those that obtain in **Three Charities**, in identical-positions precisifications of our sequence of

ten **Trolley** situations, any plausible version of IMM will imply that it is most appropriate for the decision maker to pull the lever in each of the ten **Trolley** situations.

Fortunately, this strikes me as an attractive resolution of the identical-positions precisification of our sequence of ten **Trolley** situations. Once again (recall §2.4 above), underwriting this kind of ‘hedging’ in a sequence of ten **Trolley** choice situations is an attractive implication that IMM has in common with its chief rival MEC (assuming *arguendo* that choiceworthiness units are intertheoretically comparable between T_1 and T_2 , and that a difference of one T_1 -unit is approximately equal to a difference of one T_2 -unit).

This sequence of ten **Trolley** choice situations is a sequence in which gains from contract are available to the theory representatives. However, we can also imagine sequences of discrete choice situations in which there are no available gains from trade or contract. For instance, imagine a sequence of ten identical instances of the following choice situation:

Special Obligation: some decision maker faces a choice between two possible options:

- (i) averting a lesser harm from befalling the decision maker’s parents;
- (ii) averting a greater harm from befalling a stranger.

This decision maker has 90% credence in an impartialist moral theory T_1 according to which she is required to aid the stranger,

but 10% credence in a partialist moral theory T_2 according to which she is required to aid her parents.

As we always do, let us again assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations other than this sequence of ten in which T_1 and T_2 disagree. This rules out the potential for any gains from trade or contract involving other scenarios in addition to our sequence of ten instances of **Special Obligation**. Furthermore, let us also suppose (as we did for **Trolley**) that according to both T_1 and T_2 , the choiceworthinesses of the two options available in any of our ten instances of **Special Obligation** do not depend on that instance's position in the sequence, or upon how the decision maker has behaved in any **Special Obligation** choice situations that she has already encountered.

Under these assumptions, it makes sense for IMM's initial assignment of control rights to endow R_1 with the right to choose how the decision maker will behave in nine instances of **Special Obligation** in our sequence of ten, with R_2 being endowed with this right in the sole remaining instance of **Special Obligation**.

After this initial endowment of control rights, R_1 and R_2 would then have the option to trade or make contracts with each other. As it happens, however, no mutually beneficial trades or contracts are available in these circumstances. R_1 and R_2 's preferences concerning our sequence of ten instances of **Special Obligation** are diametrically opposed, in the sense that any change in the number of choice situations in which the decision maker

aids her parents instead of the stranger that is supported by either R_1 or R_2 will always be opposed by the other theory representative. Thus, there are no opportunities for gains from trade or contract within our sequence of ten instances of **Special Obligation**. Moreover, since I have stipulated that our decision maker believes with certainty that she will never encounter any other choice situations after **Special Obligation** wherein T_1 and T_2 issue incompatible directives, R_1 and R_2 have identical preferences over how the decision maker should behave in any choice situations other than **Special Obligation**, there are also no opportunities for any other gains from trade or contract either.

Under these conditions, R_1 and R_2 will not agree to any trades or contracts in the IMM model of our sequence of ten instances of **Special Obligation**. Instead, R_1 will instruct the decision maker to aid the stranger in nine out of our ten instances of **Special Obligation**, and R_2 will instruct the decision maker to aid her parents in the one remaining instance. Hence, IMM implies that aiding the stranger nine times out of ten is the most appropriate response to our sequence of ten instances of **Special Obligation**. This case illustrates that according to IMM, in certain sequences of choice situations it can sometimes be appropriate for a morally uncertain decision maker to alternate between the different options favoured by each of the moral theories in which she has positive credence.

An alternative response to our sequence of ten instances of **Special Obligation** that the decision maker might have considered would have been for

her to aid the stranger in all ten choice situations – always choosing the option favoured by the moral theory T_1 in which she has 90% credence. However, this response strikes me as intuitively inappropriate. Always aiding the stranger would seem to cede too much influence to a moral theory T_1 about which the decision maker still has some doubts, and too little influence to a moral theory T_2 in which the decision maker has 10% credence. Affording each theory its due influence over the decision maker's choices seems to me to require her to at least sometimes aid her parents, even if she usually does not.

In my view, then, IMM supplies us with a plausible response to our sequence of ten identical instances of **Special Obligation**. In fact, IMM honours my intuitions here much more faithfully than its main rival, MEC. Depending on how we spell out the details of T_1 and T_2 's views about the choiceworthinesses of the two options available in **Special Obligation**, MEC could imply that in our sequence of ten instances of **Special Obligation**, either:

- (1) that the decision maker always aiding the stranger is the most appropriate response;
- (2) that the decision maker always aiding her parents is the most appropriate response; or
- (3) that no possible sequence of actions is any more or less appropriate than any other.¹

¹(1) is perhaps the most likely implication of MEC in these circumstances. Nonethe-

Under no circumstances, however, could MEC ever imply that it is most appropriate for the decision maker to aid the stranger in exactly nine out of ten instances of **Special Obligation**. Overall, then, IMM honours my intuitions about our sequence of ten instances of **Special Obligation** much more faithfully than its main rival does.

In fact, IMM also has this advantage over several of its other rivals as well, including MFT and MFO, since MFT and MFO both imply that it is most appropriate for our decision maker to aid the stranger in all ten instances of **Special Obligation**. This uncompromising, ‘winner takes all’ response to this sequence of choice situations once again strikes me as an important disadvantage of MFT and MFO as compared against the IMM approach (as was also the case in §2.3 above).

To summarize: if we suppose that the decision maker knows she is about to confront a suitably divisible number of identical instances of the **Trolley** or **Special Obligation** choice situations, and if we also make a few further assumptions about T_1 and T_2 choiceworthiness evaluations in sequences like this, then it is not too difficult to see how we could specify an attractive IMM response to these circumstances. IMM’s initial assignment of control rights should simply endow R_1 and R_2 each with the right to choose how the

less, MEC could instead imply (2) if we assumed that the difference in choiceworthiness between aiding one’s parents and aiding a stranger is far lower according to T_1 than it is according to T_2 . (This is a reflection of the ‘fanaticism’ of MEC.) Furthermore, MEC could instead imply (3) if the expected choiceworthiness of aiding a stranger just so happened to be exactly equal to the expected choiceworthiness of aiding one’s parents in **Special Obligation**.

decision maker will behave in some share of the situations that she is about to confront, with R_1 and R_2 's shares being proportional to the decision maker's credences in T_1 and T_2 .

5.2 Lotteries

However, this still leaves us without an answer in – for instance – cases where our decision maker (at least for all she knows) only confronts a one-off instance of some discrete-choice situation like **Trolley** or **Special Obligation**. How should IMM model these kinds of one-off discrete-choice situations?

One natural approach is to stipulate that in one-off discrete choice situations like **Trolley** or **Special Obligation**, each of the theory representatives should be modelled as being initially endowed with *tickets* in a *lottery* to determine which representative will choose the option to be selected by the decision maker.² For instance, in IMM's model of a one-off instance of **Trolley**, R_1 and R_2 would each be endowed with a lottery ticket that gives them a 50% chance of determining whether the decision maker will push the heavy-set man, pull the lever, or do nothing (since the decision maker in **Trolley** has 50% credence in each of the two moral theories T_1 and T_2). This lottery approach in effect converts the discrete, indivisible 'right to determine what the decision maker will do in **Trolley**' into a continuously divisible resource, *viz.* the *chance* to determine what the decision maker will do in **Trolley**.

²Newberry and Ord 2021; Greaves and Cotton-Barratt 2023, §4.2; Kaczmarek, Lloyd and Plant 2025.

After this initial endowment of lottery tickets, R_1 and R_2 would then have the option to trade or make contracts with each other. Are any mutually beneficial trades or contracts available to the two theory representatives under these circumstances?

Imagine, first of all, an outcome in which R_1 and R_2 do not make any trades or contracts with each other. Under these conditions, if R_1 wins the decision lottery, then it will be in R_1 's best interests to instruct the decision maker to push the heavysset man; and if R_2 wins the decision lottery, then it will be in R_2 's best interests to instruct the decision maker to do nothing. Unfortunately, neither R_1 nor R_2 is likely to be particularly happy *ex ante* with this lottery over possible outcomes, since each representative thinks that it would carry a 50% risk of the decision maker performing a highly unchoiceworthy option.

Now imagine, however, an outcome in which R_1 and R_2 agree in advance that either representative will instruct the decision maker to pull the lever if that representative wins the decision lottery. Since R_1 and R_2 both think that pulling the lever is almost as choiceworthy as their favoured option in **Trolley**, it is reasonable to assume that both theory representatives will regard the decision maker pulling the lever with certainty as a significant improvement *ex ante* over the decision maker randomising 50-50 over pushing the heavysset man and doing nothing. In other words, it makes sense to assume that randomising 50-50 over pushing and doing nothing is Pareto dominated by (among other things) pulling the lever with certainty. Since

no rational economic agents with the ability to make contracts would ever settle for a Pareto-dominated outcome, IMM thus suggests that the decision maker randomising 50-50 over pushing the heavyset man and doing nothing would be inappropriate in the one-off **Trolley** choice situation.

Of course, to repeat a by-now familiar theme, this discussion just raises the question of which option or lottery over options would be *most* appropriate in response to the one-off **Trolley** choice situation. And, as always, the answer to this question will depend on the details of how IMM should model the inter-representative bargaining process, as well as on our specification of exactly how T_1 and T_2 compare the choiceworthinesses of potential lotteries over the three options available in **Trolley**.

Fortunately, though, there are nonetheless some ways of stipulating T_1 and T_2 's views about the choiceworthiness of option lotteries for **Trolley** under which any plausible model of inter-representative IMM bargaining will imply that the most appropriate outcome in the one-off **Trolley** choice situation is for the decision maker to pull the lever with certainty. Once again, I have in mind cases in which R_1 and R_2 's bargaining positions are structurally identical to each other.

For example, suppose once again that according to T_1 , pushing the heavyset man has a choiceworthiness value of $+1 T_1$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_1$ -units, and doing nothing has a negative choiceworthiness value of $-1 T_1$ -units. Symmetrically, suppose that according to T_2 , doing nothing has a choiceworthiness value of $+1 T_2$ -units,

whereas pulling the lever has a choiceworthiness value of $+0.9 T_2$ -units, and pushing the heavysset man has a negative choiceworthiness value of $-1 T_2$ -units.

Furthermore, suppose that T_1 and T_2 both have the same *risk attitude* with respect to lotteries over options in **Trolley**.³ In particular, suppose for the sake of simplicity that T_1 and T_2 are both *risk neutral* with respect to lotteries over options in **Trolley**. In other words, according to both T_1 and T_2 , the *ex ante* choiceworthiness of any risky lottery over options in **Trolley** is simply the *expected* choiceworthiness of the decision maker's behaviour under that lottery. For instance, the *ex ante* choiceworthiness of the decision maker randomising 50-50 over pushing the heavysset man and doing nothing will be $(0.5 \times 1) + (0.5 \times -1) = 0 T_1$ -units according to T_1 , and $(0.5 \times 1) + (0.5 \times -1) = 0 T_2$ -units according to T_2 .

Under these assumptions, R_1 and R_2 's bargaining positions are structurally identical to each other in the sense defined in §2.4 above. In every respect that could plausibly make a difference to IMM's bargaining process, R_1 's bargaining position vis-à-vis R_2 is identical to R_2 's bargaining position vis-à-vis R_1 .

Under these conditions, no plausible specification of IMM's bargaining process could privilege either of these two theory representatives over the other representative in the one-off **Trolley** choice situation. For instance, no plausible version of IMM could privilege R_1 over R_2 by implying that it

³I will discuss the importance of this assumption in greater detail in §5.3 below.

is most appropriate for the decision maker to randomise 80-20 between (i) pulling the lever (the compromise option) and (ii) pushing the heavysset man (R_1 but not R_2 's favourite option). Since R_1 and R_2 's bargaining positions are structurally identical to each other, there could be no reasonable basis for privileging R_1 's objectives over R_2 's like this.

Thus, under the assumption that R_1 and R_2 's bargaining positions are structurally identical, any plausible version of IMM will imply that the most appropriate lottery over options in the one-off **Trolley** choice situations assigns no higher and no lower probability to pushing the heavysset man than it assigns to doing nothing.

Furthermore, our new assumptions about T_1 and T_2 also imply that

- (1) reducing by $n\%$ the decision maker's probability of pushing the heavysset man

at the same time as

- (2) reducing by $n\%$ the decision maker's probability of doing nothing,

and thereby also *eo ipso*

- (3) increasing by $2n\%$ the decision maker's probability of pulling the lever,

must overall be an *ex ante* strong Pareto improvement, in the sense that it will make both R_1 and R_2 happier *ex ante* than they otherwise would

have been. Thus, any lottery over options in the one-off **Trolley** choice situation that assigns the same nonzero probability to pushing the heavysset man as it assigns to doing nothing must be *ex ante* Pareto dominated by pulling the lever with certainty. And so given that no rational bargainers would ever settle for an *ex ante* Pareto dominated lottery, and given my new assumptions about T_1 and T_2 , any plausible version of IMM will imply that a response to the one-off **Trolley** situation that assigns the same *nonzero* probability to pushing the heavysset man as it assigns to doing nothing must be inappropriate.

To summarize: under the assumption of strictly identical bargaining positions, the most appropriate lottery over options in the one-off **Trolley** choice situation must assign a probability to pushing the heavysset man equal to the probability that it assigns to doing nothing; and, furthermore, both of these probabilities must be *zero*. Hence, assuming structurally identical bargaining positions, any plausible version of IMM will imply that it is most appropriate for the decision maker to pull the lever with certainty in the one-off **Trolley** choice situation.

Fortunately, this strikes me as an attractive resolution of the identical-positions precisification of the **Trolley** situation. Once again, underwriting this kind of ‘hedging’ in **Trolley** is an attractive implication that this lottery version of IMM has in common with its chief rival, MEC (assuming *arguendo* that choiceworthiness units are intertheoretically comparable between T_1 and T_2 , and that a difference of one T_1 -unit is approximately equal to a difference

of one T_2 -unit).

Thus, a version of IMM under which the theory representatives in one-off, discrete-choice situations like **Trolley** are each initially endowed with tickets in a decision lottery can sometimes allow for attractive resolutions of at least some of these kinds of discrete-choice situations.⁴

5.3 Problems with lotteries

5.3.1 Stochasticity

Unfortunately, however, this lottery proposal also suffers from a couple of important defects. One of these defects is that it runs the risk of introducing implausible stochasticities into IMM's verdicts about appropriateness.

For instance, consider a decision maker who is about to confront a one-off instance of the **Special Obligation** discrete-choice situation (first introduced in §5.1 above). Furthermore, suppose (like always) that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to this one-off instance of **Special Obligation**.

In the lottery IMM model for **Special Obligation**, the theory represen-

⁴**Trolley** is a one-off choice situation in which gains from contract are available to the theory representatives. Of course, we can also imagine one-off discrete-choice situations in which there are no available gains from trade or contract; however, I defer consideration of these kinds of choice situations until §5.3ff. below.

tative R_1 would be endowed with a lottery ticket that carries a 90% chance to determine whether this decision maker aids the stranger or her parents, and R_2 would be endowed with a lottery ticket that carries a 10% chance of this. After this initial endowment of lottery tickets, R_1 and R_2 will then have the option to trade or make contracts with each other. Are any mutually beneficial trades or contracts available to the two theory representatives under these circumstances?

We may reasonably assume that T_1 directs the decision maker to maximise the probability that she will aid the stranger in **Special Obligation**, whereas T_2 directs her to maximise the probability that she will aid her parents. Under these assumptions, R_1 and R_2 's preferences over the probability of the decision maker aiding the stranger versus her parents in **Special Obligation** are diametrically opposed, in the sense that any change in this probability that is supported by one of these theory representatives will always be opposed by the other. Thus, there are no opportunities for intratemporal gains from trade within the **Special Obligation** choice situation. Moreover, since I have stipulated that our decision maker believes with certainty that she will never encounter any other choice situations after **Special Obligation** wherein T_1 and T_2 issue incompatible directives, there are also no opportunities for intertemporal gains from trade either.

Under these conditions, R_1 and R_2 will not agree to any trades or contracts in the lottery version of the IMM model for the **Special Obligation** choice situation. Instead, the representative that wins the decision lottery

will simply choose the option recommended by the moral theory that it represents. In other words: there is a 90% chance that R_1 will win the lottery and then instruct the decision maker to aid the stranger, and a 10% chance that R_2 will win the decision lottery and then instruct the decision maker to aid her parents. Hence, the lottery version of IMM implies that it is most appropriate for the decision maker to randomise 90:10 between aiding the stranger and her parents in **Special Obligation**.

Unfortunately, this strikes me as a highly unattractive resolution of the **Special Obligation** choice situation (especially insofar as we assume that this is a high stakes choice situation). According to the lottery version of IMM, there is a 10% chance that it will be most appropriate for our decision maker to aid her parents in **Special Obligation**, despite the fact that she has 90% credence in a moral theory according to which the decision maker is morally required to aid the stranger in **Special Obligation**. This implication of the lottery IMM response to **Special Obligation** strikes me as highly implausible.⁵ Whether it is most appropriate for the decision maker to aid the stranger or her parents in a high-stakes choice situation like **Special Obligation** should not depend on the outcome of some morally arbitrary random process. A more plausible version of IMM would instead imply that aiding the stranger is the only appropriate option in the one-of **Special Obligation** choice situation.

The intuitively implausible stochasticity of the lottery version of IMM can

⁵Likewise Newberry and Ord 2021, p. 8; Kaczmarek, Lloyd and Plant 2025, §5.1.

be brought out particularly vividly by imagining that some decision maker who faces an entire lifetime's worth of (heterogeneous) discrete choice situations splits her credence 90:10 between two moral theories T_1 and T_2 whose directives are diametrically opposed to each other, and between whose theory representatives R_1 and R_2 there are no potential gains from trade or contract. According to the lottery version of IMM, there is a slim yet nonzero chance for it to be most appropriate for our decision maker to follow T_2 's directives for the entirety of her lifetime, in spite of the fact that she has 90% credence in a moral theory T_1 that is diametrically opposed to all of these directives. This implication strikes me as wildly implausible. A more plausible version of IMM would instead imply that it is most appropriate for the decision maker to follow T_1 's directives in most of the choice situations that she will confront in her lifetime, following T_2 's directives only very occasionally.

5.3.2 Risk attitudes

A second important defect of the simply lottery version of IMM is that it has the potential to make the appropriateness facts for some morally uncertain decision makers less sensitive than they plausibly should be to the directives of any *risk averse* moral theories in which those decision makers have positive credences.

Consider a one-off version of the **Trolley** choice situation in which T_1 and T_2 are symmetrical in every respect except their risk attitudes. For the sake of concreteness, let's suppose once again that according to T_1 , pushing the

heavysset man has a choiceworthiness value of $+1 T_1$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_1$ -units, and doing nothing has a negative choiceworthiness value of $-1 T_1$ -units. Symmetrically, suppose that according to T_2 , doing nothing has a choiceworthiness value of $+1 T_2$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_2$ -units, and pushing the heavysset man has a negative choiceworthiness value of $-1 T_2$ -units.

Asymmetrically, however, let us suppose that T_1 and T_2 have different risk attitudes with respect to lotteries over options in **Trolley**. In particular, suppose that although T_1 is risk neutral with respect to lotteries over options in **Trolley**, T_2 is by contrast risk averse. Thus, according to T_1 , the choiceworthiness of selecting any given risky lottery over possible outcomes must always be equal to a probability-weighted average of what the choiceworthinesses according to T_1 would be for options to determinately realise each of the possible final outcomes of the risky lottery. For example, the choiceworthiness of our decision maker randomising 50:50 between pushing the heavysset man and doing nothing must be $(0.5 \times 1) + (0.5 \times -1) = 0 T_1$ -units. By contrast, according to T_2 , the choiceworthiness of selecting any given risky lottery over possible outcomes must always be *lower* than a probability-weighted average of what the choiceworthinesses according to T_2 would be for options to determinately realise each of the possible final outcomes. For example, the choiceworthiness of our decision maker randomising 50:50 between pushing the heavysset man and doing nothing must be *lower*

than $(0.5 \times 1) + (0.5 \times -1) = 0$ T_1 -units.

(Perhaps T_2 is risk averse in **Trolley** because T_2 incorporates some version of the ‘precautionary principle.’⁶ Alternatively, perhaps T_2 implies that regardless of whether or not the decision maker actually violates the heavysset man’s right to life by pushing him in front of the trolley, subjecting him to a certain *risk* of death would in itself constitute a violation of another right – against risk imposition *per se* – held by the heavysset man.)

Under these assumptions, R_1 and R_2 ’s bargaining positions are *not* structurally identical to each other in the sense defined in §2.4 above. In one respect that could plausibly make a difference to IMM’s bargaining process (their risk attitudes), R_1 ’s bargaining position vis-à-vis R_2 is nonidentical to R_2 ’s bargaining position vis-à-vis R_1 .

In particular, R_1 ’s bargaining position is plausibly *stronger* than R_2 ’s in this asymmetrical precisification of the one-off **Trolley** choice situation. If R_1 and R_2 do not agree to any contracts in the lottery IMM model for **Trolley**, then there is a 50% chance that the decision maker will push the heavysset man, and a 50% chance that she will do nothing. As the more risk averse of the two theory representatives, R_2 will be more uneasy than R_1 with the prospect of defaulting to this risky lottery over outcomes. Thus, however we precisify IMM’s model of the bargaining process, it is reasonable to assume that R_2 will be more willing than R_1 to make concessions in the lottery IMM model for **Trolley**. R_2 will give up more of what it wants in the

⁶Rechnitzer 2020.

negotiations in order to avoid a lottery that randomises 50-50 over pushing the heavysset man and doing nothing.⁷

Hence, the lottery version of IMM is unlikely to imply that it is most appropriate for the decision maker to pull the lever with certainty in this precisification of the one-off **Trolley** choice situation. Instead, it is likely to imply that it is most appropriate for the decision maker to randomise between pushing the heavysset man and pulling the lever, with her probability of pushing the heavysset man being somewhat lower than her probability of pulling the lever.

Unfortunately, however, this strikes me as an unattractive resolution of this precisification of the one-off **Trolley** choice situation. According to this lottery version of IMM, there is some chance that it will be most appropriate to choose T_1 's favoured option in **Trolley**, but no chance at all that it will be most appropriate to choose T_2 's favoured option, despite the fact that T_1 and T_2 's evaluations of the three determinate options in **Trolley** are completely symmetrical. This implication of the lottery IMM response to **Trolley** strikes me as implausible. I can see no reason why the appropriateness of pulling the lever with certainty in **Trolley** should depend on T_1 and T_2 's risk attitudes. A more plausible version of IMM would instead imply that pulling the lever with certainty is the only appropriate option in this precisification of the

⁷In the words of Volij and Winter (2002), "that increasing risk aversion reduces a player's share in the bargaining outcome and increases that of his opponent" is "one of the results most frequently quoted in the bargaining literature," and this result appears in many "different variations including both the cooperative and non-cooperative frameworks."

one-off **Trolley** choice situation.

To be clear, I do not mean to deny here that IMM's appropriateness verdicts should be sensitive to our theory representatives' attitudes towards descriptive uncertainty in choice situations which ineliminably involve an element of descriptive uncertainty. For instance, in **Risky Philanthropy** situation introduced in §3 above, a philanthropic maker is descriptively uncertain about whether or not some new social enterprise corporation will succeed. Now under these sorts of conditions, I agree that it is highly plausible to suppose that IMM's appropriateness verdicts should be sensitive to our theory representatives' risk attitudes.

By contrast, however, I have argued in the present subsection that it is implausible to suppose that IMM's appropriateness verdicts should be sensitive to attitudes towards descriptive uncertainty even in choice situations like **Trolley** that need not involve any genuine descriptive uncertainty. Moreover, I have suggested that these arguments give us at least some reason to reject simple lotteries.

5.4 Social planning

5.4.1 Augmenting IMM

In the previous section, I argued that the lottery version of IMM suffers from a couple of important defects (*viz.* implausible stochasticity and penalising

risk aversion). Both of these two defects can be traced to the fact that the simple lottery version of IMM introduces too much risk into the IMM model, by creating a winner-takes-all lottery in discrete-choice situations. This risky lottery creates unwanted stochasticity in cases where the theory representatives cannot agree to a deal; and it also gives risk-averse theory representatives unduly weak initial bargaining positions. Thus, we should be looking for an alternative to the lottery proposal which avoids introducing this extra element of risk into the IMM model.

We could try to satisfy this desideratum by adopting a version of IMM which determinately assigns the decision right for each discrete-choice situation to one or another of the theory representatives. For instance, in a one-off instance of **Trolley**, either R_1 or R_2 could be determinately endowed with the control right to determine how the decision maker will behave in this choice situation. However, this approach will unfortunately necessarily sometimes fail to give each moral theory its due influence on the decision maker's plan of action. For instance, if R_1 is initially endowed with the control right in our one-off instance of **Trolley**, then it will be in R_1 's best interests to instruct the decision maker to push the heavysset man; and if R_2 is initially endowed with this control right, then it will be R_2 's best interests to instruct the decision maker to do nothing. Hence, neither of these two possible initial endowments would result in the decision maker being instructed to pull the lever in our one-off instance of **Trolley**. But, it strikes me as highly plausible to suppose that pulling the lever is the only appropriate option in any

identical-positions precisification of this choice situation, and that it uniquely affords due influence to both T_1 and T_2 .

To take stock: in §5.3 above, I have argued that discrete-choice decision rights should not be allocated to theory representatives *indeterminately* through lotteries. But then just now, I argued that these decision rights also should not be allocated to theory representatives *determinately* either.

At this point in the dialectic, then, it would be natural to worry that discrete-choice situations might pose an indissoluble problem for IMM. If it is implausible for discrete-choice decision rights to be allocated to theory representatives either indeterminately or determinately, then it might seem as though any possible version of IMM is doomed to handling discrete-choice situations implausibly.

However, this worry would mistakenly presume that any possible version of IMM must somehow allocate discrete-choice decision rights *only to theory representatives*. In fact, in the remainder of this chapter I will suggest that we should augment the roster of imaginary agents involved in the IMM decision model with a ‘*social planner*’ agent, whose sole objective will be to ensure that every theory representative comes as close as possible to having at least its due influence on the decision maker’s overall life plan. And whenever our decision maker encounters any discrete-choice situations, the social planner will be initially endowed with the control right to instruct the decision maker to choose any of the options available to her.

Moreover, in resource-division choice situations, the social planner also

has the power to *compel* any or all of the theory representatives into abiding by any trades or contracts that the planner chooses to impose on them. But of course, our planner will use these special powers of compulsion only to ensure that every theory representative comes as close as possible to having at least its due influence.

For an example of how this will work, we can reconsider the case of the decision maker who confronts a one-off instance of the identical-positions precisification of **Trolley**. In my preferred IMM model for this choice situation, the right to choose which option the decision maker will be instructed to perform is initially assigned to the imaginary social planner, whose sole objective is to ensure that every theory representative comes as close as possible to having at least its due influence.

How will the social planner use this decision right with which she has been endowed? Well, for the moment we may very plausibly suppose that in a one-off instance of the identical-position precisification of **Trolley**, the option of pulling the lever uniquely affords due influence to both T_1 and T_2 .⁸ Under this plausible supposition, the social planner will use her control right just to instruct the decision maker to pull the lever. Thus, my preferred version of IMM would imply that it is most appropriate for the decision-maker to pull the lever in our one-off instance of **Trolley**.

In this particular discrete-choice situation, there is an option available (pulling the lever) under which all of the moral theories that the decision

⁸I will eventually vindicate this assumption in §5.4.6 below.

maker has positive credence in would plausibly be afforded their due influence. Unfortunately, however, in many other discrete-choice situations there are no such options available. For instance, in **Special Obligation**, there are only two possible options available (aiding the stranger or aiding the parents) – and neither of these two options in themselves afford due influence to both T_1 and T_2 (the moral theories in which our decision maker has credences of 90% and 10% respectively). Aiding the stranger completely satisfies T_1 's directives in this choice situation, but it also completely violates T_2 's directives. And, on the flipside, aiding the parents completely satisfies T_2 's directives, but it also completely violates T_1 's directives. Hence, it would be impossible for our social planner to issue an instruction in **Special Obligation** which would in itself afford proportional influence to both T_1 and T_2 .

On the other hand, in at least some cases wherein our decision maker knows that a one-off instance of **Special Obligation** will be followed by other choice situations in which T_1 and T_2 issue incompatible directives, it might nonetheless be possible for the social planner to compel R_1 and R_2 to abide by an *intertemporal contract* which would give both theory representatives their due influence on the decision maker's overall life plan. For example, imagine that our decision maker confronts a one-off instance of **Special Obligation**, yet knows that immediately after she resolves this discrete-choice situation, she will then confront a resource-division choice situation 'X,' in which (just as in **Philanthropy**) she will need to decide how to allocate some money. Moreover, suppose that T_1 and T_2 issue diametrically

opposed directives concerning how our decision maker should allocate her money in **X**. And, as we always do, let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contracts involving other scenarios in addition to these two choice situations. Finally, let us also assume that choiceworthiness returns to scale from each of the possible uses for money in the choice situation **X** are *constant* according to both of the moral theories T_1 and T_2 .

Under this set of circumstances, the social planner might be able to mandate an intertemporal contract giving both R_1 and R_2 their due influence on our decision maker's plan of action in **Special Obligation-then-X**. One possibility here would begin with an instruction to aid the stranger in **Special Obligation**. Recall, however, that this instruction would give T_1 slightly more – and T_2 slightly less – than its due influence in this choice situation. Hence, in order for T_1 and T_2 to both have due influence on the decision maker's *overall* life plan, R_1 must be compelled to transfer to R_2 some of R_1 's influence over **X**. Thus, our social planner would want to compel R_1 to transfer to R_2 a modest tranche of R_1 's initial endowment of control rights over resources in **X**, as a kind of forced 'compensation' for the decision maker having aided the stranger in **Special Obligation**.

An alternative possible response to **Special Obligation-then-X** would instead begin with an instruction to aid the parents in **Special Obligation**. Recall, however, that this instruction would give T_1 much less – and T_2 much

more – than its due influence in this choice situation. Hence, in order for T_1 and T_2 to both have due influence on the decision maker's overall life plan, R_2 must be compelled to transfer to R_1 a large tranche of R_2 's initial endowment of control rights over resources in \mathbf{X} , as compensation for the decision maker having aided her parents in **Special Obligation**.

5.4.2 The sequence criterion

This discussion raises two important questions: (1) which of these two instructions should the social planner issue in **Special Obligation**?; and (2) exactly what size of compensation transfer should the social planner require between R_1 and R_2 after that in \mathbf{X} ?

In what follows, I will try to answer these two questions by developing a principled criterion for deciding precisely which course of action would give both T_1 and T_2 their due influence in **Special Obligation-then- \mathbf{X}** . I will call this my 'SEQUENCE CRITERION' for due influence.

Just as in §5.1 above, I will for the moment assume that according to every moral theory in which our decision maker has positive credence, the choiceworthiness of the options available in any given choice situation do not depend upon the sequence of choice situations that this decision maker has previously confronted, or upon how the decision maker has already behaved in any of these choice situations. However, I will demonstrate in §5.5.1 below how the SEQUENCE CRITERION can be elegantly generalised once we relax this initial simplifying assumption.

In any case, under this initial assumption, my SEQUENCE CRITERION for due influence is motivated by the following three claims:

- (1) the REPETITION CLAIM: some course of action ϕ in response to a one-off instance of any given set of circumstances Φ would perfectly satisfy the due influence desideratum iff repeating ϕ many times over in response to a repeated sequence of instances of Φ would also perfectly satisfy the due influence desideratum;
- (2) the PROPORTIONALITY CLAIM: in a world in which any given set of circumstances Φ is repeated many times over, splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories would induce a course of action that perfectly satisfies the due influence desideratum; and
- (3) the INDIFFERENCE CLAIM: for any given set of circumstances Ψ , IF (a) some course of action ψ_1 would perfectly satisfy the due influence desideratum in Ψ , THEN: (b) any given course of action ψ_2 will perfectly satisfy the due influence criterion in Ψ iff (c) every theory representative would be indifferent between ψ_1 and ψ_2 .

I will now discuss and defend each of these three claims.

First of all, the REPETITION CLAIM strikes me as highly *prima facie*

plausible. If some course of action gives each moral theory exactly its due level of influence, then surely repeating this course of action over and over again must also give each moral theory its due level of influence? (At least granting my temporary simplifying assumptions.) And the same goes *mutatis mutandis* for the reverse direction of REPETITION biconditional.

Now, although this REPETITION CLAIM might at first seem quite innocuous, it nonetheless turns out to be rather useful, because it allows us to in some sense *transform* between – on the one hand – questions and claims about due influence in *one-off* discrete-choice situations, and – on the other hand – questions and claims about due influence in *sequences* of discrete-choice situations. For example, rather than asking which courses of action would give each moral theory its due influence in a one-off instance of **Special Obligation-then-X**, I can now instead ask which courses of action would give each moral theory its due influence in a world where **Special Obligation-then-X** is repeated ten times over.

In §5.1 above, I have already considered the example of a world in which **Special Obligation** is repeated ten times over. In that discussion, I argued that the due influence desideratum would be perfectly satisfied if R_1 were to be endowed with the right to choose how the decision maker will behave in nine of ten instances of **Special Obligation**, leaving R_2 to be endowed with this decision right in the sole remaining instance of **Special Obligation**.

This approach can be easily extended to handle the case in which **Special Obligation-then-X** is repeated ten times over. Recall that **X** is a resource-

<i>choice situation:</i>	Sp. Obl. 1	X 1	Sp. Obl. 2	X 2	...	Sp. Obl. 9	X 9	Sp. Obl. 10	X 10
<i>initial assignment of control rights:</i>	R_1	divided 90-10	R_1	divided 90-10	...	R_1	divided 90-10	R_2	divided 90-10

Figure 5.2: Proportional division in a sequence of ten instances of **Special Obligation-then-X**

division choice situation (just like **Philanthropy**). Hence, in each of our ten repetitions of **X**, the due influence desideratum would be perfectly satisfied if control rights over the resource in **X** were to be initially divided 90:10 between R_1 and R_2 (in proportion to credences). And so – as the PROPORTIONALITY CLAIM suggests – in a world in which **Special Obligation-then-X** is repeated ten times over, the due influence desideratum would be perfectly satisfied if decision rights over this sequence of choice situations were to be divided 90:10 between R_1 and R_2 , as illustrated in figure 5.2.

Once again, although this implication of the PROPORTIONALITY CLAIM might at first seem quite innocuous, it nonetheless turns out to be rather useful, because we can leverage it (together with other auxiliary premises) to determine how other alternative possible responses to our sequence of ten instances of **Special Obligation-then-X** would stand with respect to the due influence desideratum. For instance, imagine that for some alternative possible distribution ρ_1 of control rights over this sequence of choice situations, R_1 prefers ρ_1 over the 90:10 distribution, whereas R_2 disprefers it. Then, under

these assumptions, we can plausibly infer that ρ_1 would give R_1 *greater* than its due influence over this sequence of choice situations, and R_2 *less* than its due influence. Of course, we can also imagine that for another alternative possible distribution ρ_2 of control rights over this sequence of choice situations, R_1 disprefers ρ_2 over the 90:10 distribution, whereas R_2 prefers it. Then, under these assumptions, we can plausibly infer that ρ_2 would give R_1 less than its due influence over this sequence of choice situation, and R_2 more than its due influence.

More generally, it strikes me as highly *prima facie* plausible to suppose – as the INDIFFERENCE CLAIM implies – that R_1 and R_2 will both be indifferent between the 90:10 distribution and some alternative course of action ψ in this sequence of choice situations iff ψ likewise perfectly satisfies the due influence desideratum. After all, surely ψ would give one of the theory representatives less than its due influence iff that representative prefers the 90:10 distribution over ψ . And, similarly, surely ψ would give one of the theory representatives more than its due influence iff that representative prefers ψ over the 90:10 distribution. Moreover, these considerations can be readily generalised to support my INDIFFERENCE CLAIM.

Taking the REPETITION, PROPORTIONALITY and INDIFFERENCE claims together, we can deduce my SEQUENCE CRITERION for due influence.⁹ According to the SEQUENCE CRITERION:

⁹The argument is simple enough once we set $\Psi :=$ some sequence of repeated instances of Ψ .

any given course of action ϕ in response to a one-off instance of any given set of circumstances Φ would perfectly satisfy the due influence desideratum IFF every theory representative would be indifferent between (a) repeating ϕ many times over in response to a repeated sequence of instances of Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories.

In what follows, I will now illustrate how to apply this criterion.

5.4.3 Simple compensation

Before I illustrate how to apply the SEQUENCE CRITERION to **Special Obligation-then-X**, I will first make two more temporary simplifying assumptions. Firstly, I shall assume that the relative stakes of each of **Special Obligation** and **X** are identical across the two moral theories T_1 and T_2 . (Thus, I hereby assume that no gains from trade or contract are available in any sequence of instances of **Special Obligation-then-X**.) Secondly, I shall also assume that our decision maker has to decide how to allocate a *very large* sum of money in the choice situation **X**. Although I will relax both of these two assumptions later (in §§5.4.4-5.4.5 below), for the moment they will help to simplify my exposition of the SEQUENCE CRITERION.

According to the first of these two assumptions, no gains from trade or

<i>choice situation:</i>	Sp. Obl. 1	X 1	...	Sp. Obl. 9	X 9	Sp. Obl. 10	X 10
<i>outcome:</i>	aid the stranger	divided 90-10	...	aid the stranger	divided 90-10	aid the parents	divided 90-10

Figure 5.3: 90:10 division between T_1 and T_2 's favoured options

contract are available in any sequence of instances of **Special Obligation-then-X**. Hence, in any repeated sequences of this sort, R_1 and R_2 will both use all of their endowments of control rights in the manners recommended by T_1 and T_2 respectively. For instance, in the world where **Special Obligation-then-X** is repeated ten times over and the decision rights for this sequence are split 90:10 between R_1 and R_2 , R_1 will simply instruct our decision maker to aid the stranger in nine of the ten instances of **Special Obligation**; and R_2 will instruct the decision maker to aid her parents in the single remaining instance of this choice situation. Moreover, in all ten instances of **X**, R_1 will instruct the decision maker to use 90% of her resources in the manner recommended by T_1 , and R_2 will instruct her to use the remaining 10% in the manner recommended by T_2 . This outcome is illustrated in figure 5.3.

With this result onboard, I can now illustrate how to apply my new SEQUENCE CRITERION to our one-off instance of **Special Obligation-then-X**.¹⁰ First of all, imagine that the social planner is contemplating some course

¹⁰The SEQUENCE CRITERION can and should still be applied in cases where gains from trade or contract would be available in a repeated sequence of instances of the original set of circumstances Φ . It is just that in cases like this, we can apply the SEQUENCE CRITERION for due influence only after having first calculated the gains from trade or contract that would be realised in a world where the decision rights over a repeated sequence of instances of Φ are divided between theory representatives in proportion to the decision maker's credences in the corresponding moral theories. I avoid discussing cases

<i>choice situation:</i>	Sp. Obl. 1	X 1	...	Sp. Obl. 10	X 10
<i>repeated ϕ_S:</i>	aid the stranger	endowments: 90%-\$ S to R_1 10%+\$ S to R_2	...	aid the stranger	endowments: 90%-\$ S to R_1 10%+\$ S to R_2

Figure 5.4: Repeating ϕ_S ten times over

of action ϕ_S under which: (1) the planner would instruct our decision maker to aid the stranger in **Special Obligation**; and (2) R_1 would then be compelled to transfer to R_2 control over $\$S$ units of resources in **X**. Under this course of action, the social planner will want to select a compensation amount $\$S$ such that ϕ_S will satisfy the due influence desideratum. And according to my SEQUENCE CRITERION, ϕ_S will perfectly satisfy this desideratum IF R_1 and R_2 would both be indifferent between (a) repeating ϕ_S ten times over in response to a sequence of ten instances of **Special Obligation-then-X**, as illustrated in figure 5.4, and (b) splitting the decision rights over this same sequence of choice situations 90:10 between R_1 and R_2 (illustrated in figure 5.3).

ϕ_S could perfectly satisfy this SEQUENCE CRITERION iff it required a relatively *modest* $\$S$ compensation transfer from R_1 to R_2 in **X**. If no compensation were to change hands under ϕ_S , then R_1 would somewhat prefer the repeated ϕ_S course of action over the 90:10 division, whereas R_2 would somewhat disprefer it. After all, recall that according to both T_1 and T_2 , the choiceworthiness of performing any sequence of actions is equal to the sum of all the choiceworthiness of the sequence's constituent actions considered

like this only for ease of exposition.

in isolation. Hence, all else being equal, a world in which the decision maker aids the stranger nine times and aids her parents only once is somewhat worse according to R_1 – although somewhat better according to R_2 – than a world in which the decision maker aids the stranger ten times and her parents not at all. Moreover, recall that the relative stakes of **Special Obligation** and **X** are identical across each of the two moral theories T_1 and T_2 . Hence, in order to render both R_1 and R_2 indifferent between the repeated ϕ_S course of action and the 90:10 division, ϕ_S would have to require a relatively *modest* $\$_S$ compensation transfer from R_1 to R_2 in **X**.

Hence, under my simplifying assumptions, in a one-off instance of **Special Obligation-then-X**, if the social planner follows the SEQUENCE CRITERION for due influence, and instructs our decision maker to aid the stranger in **Special Obligation**, then R_1 should afterwards be required to transfer to R_2 a modest share of R_1 's initial endowment of control rights in the subsequent resource-division choice situation **X**. Fortunately, this strikes me as a plausible implication. After all, our decision maker is not 100% certain in the moral theory T_1 according to which she should aid the stranger in **Special Obligation**, since this decision maker has 10% credence in the moral theory T_2 according to which she should aid her parents instead. But, aiding the stranger in **Special Obligation** corresponds to T_1 having *total* control over the decision maker's behaviour in this choice situation. Hence, in order for T_1 to have no more than its due influence on the decision maker's *overall* life plan, T_1 needs to be given slightly less influence than it otherwise would

<i>choice situation:</i>	Sp. Obl. 1	X 1	...	Sp. Obl. 10	X 10
<i>repeated ϕ_P:</i>	aid the parents	endowments: 90%+\$ P to R_1 10%-\$ P to R_2	...	aid the parents	endowments: 90%+\$ P to R_1 10%-\$ P to R_2

Figure 5.5: Repeating ϕ_P ten times over

have over the decision maker's behaviour in **X**. And, of course, the same goes *mutatis mutandis* for T_2 . Thus, the SEQUENCE CRITERION strikes me as having plausible implications in this set of circumstances.

I now turn to considering what the SEQUENCE CRITERION would imply in our one-off instance of **Special Obligation-then-X** if the social planner were instead contemplating some course of action ϕ_P under which: (1) the social planner would instruct our decision maker to aid her parents in **Special Obligation**; and (2) R_2 would then be compelled to transfer to R_1 control over $\$P$ units of resources in **X**. Once again, our social planner will want to select a compensation amount $\$P$ such that ϕ_P will satisfy the due influence desideratum. And according to my SEQUENCE CRITERION, ϕ_P will perfectly satisfy this desideratum IF R_1 and R_2 would both be indifferent between (a) repeating ϕ_P ten times over in response to a sequence of ten instances of **Special Obligation-then-X**, as illustrated in figure 5.5, and (b) splitting the decision rights over this same sequence of choice situations 90:10 between R_1 and R_2 (as illustrated in figure 5.3).

ϕ_P could perfectly satisfy this SEQUENCE CRITERION iff it required a relatively *large* $\$P$ compensation transfer from R_2 to R_1 in **X**. If no compensation were to change hands under ϕ_P , then R_1 would greatly disprefer the repeated

ϕ_P course of action over the 90:10 division, whereas R_1 would greatly prefer it. After all, all else being equal a world in which the decision maker aids the stranger nine times and aids her parents only once is much better according to R_1 – although much worse according to R_2 – than a world in which the decision maker aids her parents ten times and the stranger not at all. Furthermore, recall once again that the relative stakes of **Special Obligation** and **X** are identical across each of the two moral theories T_1 and T_2 . Hence, in order to render both R_1 and R_2 indifferent between the repeated ϕ_P course of action and the 90:10 division, ϕ_P would have to require a relatively *large* $\$P$ compensation transfer from R_2 to R_1 to **X**.

Hence, under my simplifying assumptions, in a one-off instance of **Special Obligation-then-X**, if the social planner follows the SEQUENCE CRITERION for due influence, and instructs our decision maker to aid her parents in **Special Obligation**, then R_2 should afterwards be required to transfer to R_1 a large share of R_2 's initial endowment of control rights in the subsequent resource-division choice situation **X**. Fortunately, this strikes me as a plausible implication. After all, our decision maker only has 10% credence in the moral theory T_2 according to which she should aid her parents, whereas she has 90% credence in the moral theory T_1 according to which she should not. Hence, the decision maker aiding her parents in **Special Obligation** corresponds to T_1 having much less than its due influence in this choice situation. Thus, in order for T_1 to have its due influence on the decision maker's *overall* life plan, T_1 needs to be given significantly greater influence than it otherwise

would have over the decision maker's behaviour in **X**. And, of course, the same goes *mutatis mutandis* for T_2 . Therefore, the SEQUENCE CRITERION has plausible implications in this case.

5.4.4 Compensatory scope

In the previous subsection, I demonstrated that for at least some precisifications of **Special Obligation-then-X**, satisfying the SEQUENCE CRITERION is perfectly compatible with *either* of the two possible instructions which the social planner could issue to the decision maker in **Special Obligation**. However, this result was in fact an artifact of my temporary simplifying assumptions that: (i) the relative stakes of **Special Obligation** and **X** are identical across each of the two moral theories T_1 and T_2 ; and (ii) the decision maker has to decide how to allocate a *very large* sum of money in the choice situation **X**.

In what follows, I will discuss how the social planner should handle precisifications of **Special Obligation-then-X** that violate one or both of my simplifying assumptions. I will begin (in the present subsection) by relaxing only the assumption that **X** involves a very large sum of money. Thus, I will continue to assume for the moment that the relative stakes of **Special Obligation** and **X** are each identical across T_1 and T_2 . However, I will relax this assumption too in §5.4.5 below.

Recall from §5.4.3 above that (under the assumption of identical relative stakes):

(i) ϕ_S could perfectly satisfy the SEQUENCE CRITERION iff it involved a *modest* $\$S$ compensation transfer from R_1 to R_2 in \mathbf{X} ; and

(ii) ϕ_P could perfectly satisfy the SEQUENCE CRITERION iff it involved a *large* $\$P$ compensation transfer from R_2 to R_1 in \mathbf{X} .

Thus, if the amount of money available in \mathbf{X} affords my social planner ample scope for compensation, then satisfying the SEQUENCE CRITERION is perfectly compatible with both possible instructions that the social planner could issue to the decision maker in **Special Obligation**.

However, in precisifications of **Special Obligation-then- \mathbf{X}** wherein \mathbf{X} affords a more limited scope for compensation, one or both of these two instructions could be incompatible with perfectly satisfying the SEQUENCE CRITERION. For instance, if R_2 's initial 10% endowment of control rights in \mathbf{X} were to be less than the large $\$P$ compensation transfer required for due influence in ϕ_P , then an instruction to aid the parents in **Special Obligation** would be incompatible with perfectly satisfying the SEQUENCE CRITERION. Even if R_2 were required to transfer to R_1 *all* of R_2 's initial 10% endowment of control rights in \mathbf{X} , this might nonetheless fall short of the compensation transfer required to give each moral theory its due influence under ϕ_P .

What about ϕ_S ? Well, under these circumstances, R_1 's initial 90% endowment of control rights in \mathbf{X} *might or might not* still be larger than the modest $\$S$ compensation required for due influence in ϕ_S . If a *medium* sum of money were to be available in \mathbf{X} , then ϕ_S but not ϕ_P would be perfectly

compatible with satisfying the SEQUENCE CRITERION. Hence, in this precisification of **Special Obligation-then-X**, our social planner can perfectly satisfy the SEQUENCE CRITERION only if she (1) instructs the decision maker to aid the stranger in **Special Obligation**, and then (2) compels R_1 to transfer to R_2 control over a modest tranche of resources in **X**. Thus, my preferred version of IMM implies that the most appropriate response to this particular version of **Special Obligation-then-X** will involve our decision maker (1) aiding the stranger in **Special Obligation**, and then (2) allocating slightly less than 90% of her total monetary endowment in **X** to T_1 's favoured use for it, with the remainder (slightly more than 10%) being allocated to T_2 's favoured use. After all, only this course of action gives each moral theory its due influence on the decision maker's overall life plan.

On the other hand, if only a *very small* sum of money were to be available in **X**, then neither ϕ_P nor ϕ_S would be perfectly compatible with satisfying the SEQUENCE CRITERION in **Special Obligation-then-X**. Even if R_1 were required to transfer to R_2 *all* of R_1 's initial 90% endowment of control rights in **X**, this would nonetheless fall short of the compensation transfer required to give each moral theory its due influence under ϕ_S . Hence, in this precisification of **Special Obligation-then-X**, it will be completely *impossible* for our social planner to perfectly satisfy the SEQUENCE CRITERION.

Which course of action should the social planner select in sets of circumstances like this wherein it is impossible to perfectly satisfy the SEQUENCE CRITERION? Well, in any set of circumstances like this, it strikes me as *prima*

facie plausible to suppose that our social planner should select the course of action that would *minimise the shortfall* from giving all of the theory representatives at least their due influence.

Any given version of the ϕ_S response to a one-off instance of **Special Obligation-then-X** would come close to giving every theory representative at least its due influence to the extent that it would come close to rendering R_1 and R_2 at worst indifferent between (a) repeating ϕ_S ten times over in response to a sequence of ten instances of **Special Obligation-then-X** (as illustrated in figure 5.4), and (b) splitting the decision rights over this same sequence of choice situations 90:10 between R_1 and R_2 (as illustrated in figure 5.3). Thus, if the choice situation **X** involved only a very small sum of money, then the version of ϕ_S that would minimise the shortfall from giving every theory representative at least its due influence would be the version of ϕ_S that incorporates the largest feasible $\$S$ compensation transfer from R_1 to R_2 in **X**.

Similarly, any given version of the ϕ_P response to a one-off instance of **Special Obligation-then-X** would come close to giving every theory representative at least its due influence to the extent that it would come close to rendering R_1 and R_2 at worst indifferent between (a) repeating ϕ_P ten times over in response to a sequence of ten instances of **Special Obligation-then-X** (as illustrated in figure 5.5), and (b) splitting the decision rights over this same sequence of choice situations 90:10 between R_1 and R_2 (as illustrated in figure 5.3). Thus, if the choice situation **X** involved only a very small sum of

money, then the version of ϕ_P that would minimise the shortfall giving every theory representative at least its due influence would be the version of ϕ_P that (once again) incorporates the largest feasible $\$P$ compensation transfer from R_2 to R_1 in \mathbf{X} .

Furthermore, in this one-off instance of **Special Obligation-then-X** wherein \mathbf{X} involves a very small sum of money, there is clearly some intuitive sense in which the version of ϕ_S with maximal $\$S$ compensation would come much closer to giving every theory representative at least its due influence than would the version of ϕ_P with maximal $\$P$ compensation. Thus, it strikes me as plausible to suppose that the social planner should (1) instruct our decision maker to aid the stranger in **Special Obligation**, and then should (2) compel R_1 to transfer to R_2 all of R_1 's endowment of control rights over resources in \mathbf{X} . Hence, my preferred version of IMM implies that the most appropriate response to this version of **Special Obligation-then-X** would involve our decision maker (1) aiding the stranger in **Special Obligation**, and then (2) allocating 100% of her total monetary endowment in \mathbf{X} to T_2 's favoured use for it.

Once again, this strikes me as a plausible implication. If the amount of money at stake in \mathbf{X} is inconsequential relative to what is at stake in **Special Obligation**, then it is plausibly most appropriate for our decision maker to defer to the moral theory T_1 in which she has 90% credence in deciding how to behave in **Special Obligation**. After that, however, our decision maker should then ensure that T_2 has at least some influence on the decision maker's

overall life plan by deferring to T_2 in deciding how to allocate her resources in \mathbf{X} . Thus, in this set of circumstances it is plausible to stipulate that the social planner should select the course of action that minimises the shortfall from giving every theory representative at least its due influence.

In this particular one-off instance of **Special Obligation-then- \mathbf{X}** , we can rely on our intuitions to determine which possible course of action minimises the shortfall from giving every theory representative at least its due influence. However, in some other more complicated sets of circumstances, it might not be intuitively obvious which possible course of action would minimise this shortfall. In order to handle these cases, we will need to give a precise explication of how our social planner should measure the extent to which any given course of action ϕ falls short of giving every theory representative at least its due influence (in any given set of circumstances Φ).

Fortunately, I have already discussed the topic of measuring overall shortfalls from indifference in §§4.2.3-4.2.4 above. I will not revisit those discussions here, except to note that the methods for measuring shortfalls which I developed in those earlier discussions are all readily applicable to sequences involving discrete-choice situations.

5.4.5 Relative stakes

In the previous subsection, I discussed how our social planner should handle one-off instances of **Special Obligation-then- \mathbf{X}** wherein \mathbf{X} affords only limited scope for compensation. In that discussion, I continued to assume

for ease of exposition that the relative stakes of **Special Obligation** and **X** are each identical across T_1 and T_2 . But in the rest of this subsection, I will now discuss how the social planner should handle precisifications of **Special Obligation-then-X** that violate this equal-stakes assumption.

For the sake of concreteness, I will begin by considering precisifications of **Special Obligation-then-X** wherein – relative to **X** – the stakes in **Special Obligation** are higher according to T_2 than they are according to T_1 . Perhaps according to T_1 total choiceworthiness is not particularly sensitive to who the decision maker aids in **Special Obligation**, but is highly sensitive to how she allocates even small amounts of money in **X**. And on the other hand, perhaps according to T_2 total choiceworthiness in **Special Obligation-then-X** is highly sensitive to who the decision maker aids in **Special Obligation**, but is not particularly sensitive to how she allocates even large amounts of money in **X**.

Under this new assumption about relative stakes, an instruction to aid the stranger in **Special Obligation** would be incompatible with perfectly satisfying the SEQUENCE CRITERION in **Special Obligation-then-X**. This is because it will be impossible for any ϕ_S compensation transfer in ϕ_S to render *both* R_1 and R_2 indifferent between (a) repeating ϕ_S ten times over in response to a sequence of ten instances of **Special Obligation-then-X** (as illustrated in figure 5.4), and (b) splitting the decision rights over this same sequence of choice situations 90:10 between R_1 and R_2 (as illustrated in figure 5.2). Because R_1 cares about each instance of **Special Obligation**

much less than she cares about even small amounts of money in \mathbf{X} , ϕ_S could render R_1 indifferent between the repeated ϕ_S course of action and the 90:10 division only if it required R_1 to transfer a *relatively modest* amount $\$S$ of compensation to R_2 in \mathbf{X} . However, R_2 by contrast cares about each instance of **Special Obligation** much more than she cares about even large amounts of money in \mathbf{X} ; and so ϕ_S could render R_2 indifferent between these two courses of action only if it required R_1 to transfer a *relatively large* amount $\$S$ of compensation to R_2 in \mathbf{X} . Thus, it is impossible for any version of ϕ_S to perfectly satisfy the SEQUENCE CRITERION under this set of assumptions. Because no amount of compensation can be both ‘large’ and ‘modest’ simultaneously, thus there cannot exist any possible $\$S$ transfers that would render both R_1 and R_2 indifferent.

However, under these assumptions an instruction to aid the *parents* in **Special Obligation** could still be perfectly compatible with satisfying the SEQUENCE CRITERION – provided that the amount of money available in \mathbf{X} affords sufficient scope for compensation. Under these conditions, there will exist some possible $\$P$ compensation transfer which could render both R_1 and R_2 indifferent between (a) repeating ϕ_P ten times over, and (b) the proportional split for this sequence. Hence, my preferred version of IMM would imply that it is most appropriate for our decision maker to (1) aid the parents in **Special Obligation**, and then (2) allocate more than 90% of her total monetary endowment in \mathbf{X} to T_1 ’s favoured use for it, with the remainder (less than 10%) being allocated to T_2 ’s favoured use.

Once again, this strikes me as a plausible implication. If – relative to **X** – the stakes in **Special Obligation** are higher according to T_2 than they are according to T_1 , then it is plausibly most appropriate for our decision maker to defer to T_2 in deciding how to behave in **Special Obligation**. After that, however, our decision maker should then ensure that each moral theory gets its due influence on the decision maker’s overall life plan by deferring to T_1 more – and to T_2 less – than would have otherwise been appropriate for her in the choice situation **X**. Thus, the SEQUENCE CRITERION has plausible implications in this case.

What about the set of subcases wherein (i) ϕ_S continues to be incompatible with perfectly SEQUENCE CRITERION, because – relative to **X** – the stakes in **Special Obligation** are higher according to T_2 than they are according to T_1 , but (ii) ϕ_P is also incompatible with perfectly satisfying the SEQUENCE CRITERION, because the amount of money available in **X** affords insufficient scope for compensation. Well, in these precisifications of **Special Obligation-then-X**, it will be completely impossible for our social planner to perfectly satisfy the SEQUENCE CRITERION. And I have already suggested that in sets of circumstances like this, it is plausible to stipulate that our social planner should select the course of action that minimises the shortfall from perfectly satisfying this CRITERION (recall §5.4.4 above).

In order to determine which course of action would come closest to perfectly satisfying this CRITERION in any given precisification from this set of subcases, we would need to learn the details of that particular precisification.

For one thing, we would need to know exactly to what extent, relative to \mathbf{X} , the stakes in **Special Obligation** are higher according to T_2 than they are according to T_1 . Furthermore, we would also need to learn exactly to what extent the scope for compensation under \mathbf{X} falls short of the minimum scope under which some version of ϕ_P could perfectly satisfy the SEQUENCE CRITERION.

In some possible precisifications of **Special Obligation-then- \mathbf{X}** , ϕ_P would come closer than ϕ_S to satisfying the SEQUENCE CRITERION. For instance, this result will obtain in any precisifications wherein (i) the scope for compensation in \mathbf{X} does not fall far short of the minimum scope for compensation under which ϕ_P could perfectly satisfy the SEQUENCE CRITERION, and at the same time (ii) relative to \mathbf{X} , the stakes in **Special Obligation** according to T_1 are *much* lower than they are according to T_2 .

However, in some other possible precisifications of **Special Obligation-then- \mathbf{X}** , ϕ_S would come closer than ϕ_P to satisfying the SEQUENCE CRITERION. For instance, this result will obtain in any precisifications wherein (i) the scope for compensation in \mathbf{X} falls very far short of the minimum scope for compensation under which ϕ_P could perfectly satisfy the SEQUENCE CRITERION, and at the same time (ii) relative to \mathbf{X} , the stakes in **Special Obligation** according to T_1 are *only slightly* lower than they are according to T_2 .

In general, though, it strikes me as plausible to suppose that under these circumstances, our social planner should select the course of action that

would come closest to satisfying the SEQUENCE CRITERION. Hence, under these conditions my preferred version of IMM will imply that the relative appropriatenesses of the ϕ_P and ϕ_S responses to **Special Obligation-then-X** should be determined by (i) the extent to which the scope for compensation in **X** falls short of the minimum scope under which ϕ_P could perfectly satisfy the SEQUENCE CRITERION, and (ii) the extent to which, relative to **X**, the stakes in **Special Obligation** according to T_1 are lower than they are according to T_2 .

Once again, this strikes me as a plausible implication. On the one hand, if the amount of money at stake in **X** is inconsequential relative to the amount of choiceworthiness at stake in **Special Obligation**, then this is one factor that intuitively speaks in favour of our decision maker deferring in **Special Obligation** to the moral theory T_1 in which she has 90% credence. However, on the other hand, if – relative to **X** – the stakes in **Special Obligation** according to T_1 are lower than they are according to T_2 , then this is a factor that intuitively speaks in favour of our decision maker deferring in **Special Obligation** to the moral theory T_2 according to which the relative stakes in this choice situation are highest. Overall, then, the relative appropriatenesses of ϕ_P and ϕ_S under these circumstances will plausibly depend upon which of these two factors is more decisive in any particular version of **Special Obligation-then-X**.

This concludes my discussion of the set of cases in which – relative to **X** – the stakes in **Special Obligation** are lower according to T_1 than they

are according to T_2 . But what about the case in which the relative stakes in **Special Obligation** are *higher* according to T_1 than they are according to T_2 ? Well, for the sake of brevity, I will not discuss this set of cases in detail. Instead, I will simply report that under these conditions, it will be ϕ_P that cannot possibly perfectly satisfy the SEQUENCE CRITERION. Hence, some version of ϕ_S will be the possible course of action that minimises the shortfall from perfectly satisfying this CRITERION. (And if \mathbf{X} affords sufficient scope for compensation, then ϕ_S could in fact be *perfectly* compatible with satisfying the SEQUENCE CRITERION.) Yet again, this strikes me as a plausible implication.

5.4.6 Discrete-choice sets

Thus far in my discussion of social planning, I have focused on sets of circumstances wherein the decision maker knows that a one-off instance of **Special Obligation** will be followed by a one-off instance of the resource-division choice situation \mathbf{X} in which T_1 and T_2 issue incompatible directives. By design, then, I have been considering sets of circumstances wherein R_1 and R_2 can receive compensation in the form of control rights over a continuously divisible resource. However, my social planning approach is also designed to handle sets of circumstances in which the theory representatives can receive compensation only in the form of control rights over future discrete-choice situations.

For instance, imagine that our decision maker confronts a one-off instance

of **Special Obligation**, but knows that immediately after she resolves this choice situation, she will then confront another discrete-choice situation **Y**. Furthermore, suppose that T_1 and T_2 disagree about which option the decision maker should choose in **Y**. Finally, as we always do, let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to these two choice situations.

In my preferred IMM model for this set of circumstances (**Special Obligation-then-Y**), the social planner will be initially endowed with control rights over our decision maker's behaviour in both **Special Obligation** and **Y**. As always, the social planner's sole objective will be to ensure that the theory representatives come as close as possible to having at least their due influences – as assessed by the SEQUENCE CRITERION for this desideratum. According to the SEQUENCE CRITERION, and given course of action ϕ in response to a one-off instance of **Special Obligation-then-Y** would perfectly satisfy the due influence desideratum IFF every theory representative would be indifferent between (a) repeating ϕ ten times over in response to a sequence of ten instances of **Special Obligation-then-Y**, and (b) splitting the decision rights over this sequence of choice situations 90:10 between R_1 and R_2 , as illustrated in figure 5.6.

It is unlikely that any course of action in **Special Obligation-then-Y** will perfectly satisfy this CRITERION. Since **Special Obligation** and **Y** are

<i>choice situation:</i>	Sp. Obl. 1	Y 1	Sp. Obl. 2	Y 2	...	Sp. Obl. 9	Y 9	Sp. Obl. 10	Y 10
<i>initial assignment of control rights:</i>	R_1	R_1	R_1	R_1	...	R_1	R_1	R_2	R_2

Figure 5.6: Proportional division in a sequence of ten instances of **Special Obligation-then-Y**.

both discrete-choice situations, our social planner will only have a finite menu of possible courses of action available to her in **Special Obligation-then-Y**. And we have no reason to suppose that one of these courses of action will just so happen to perfectly satisfy the SEQUENCE CRITERION for due influence.

Fortunately, it again strikes me as plausible to stipulate that in cases like this, our social planner should simply select the course of action that *minimises the shortfall* from perfectly satisfying the SEQUENCE CRITERION (recall §5.4.4 above). Suppose, for instance, that T_1 and T_2 both imply that the stakes in **Y** are fairly low relative to those in **Special Obligation**. Then under these conditions, the SEQUENCE CRITERION would probably come closest to being satisfied by a course of action under which: (1) the social planner would instruct our decision maker to aid the stranger (as favoured by R_1) in **Special Obligation**; and then (2) the social planner would instruct the decision maker to choose an option favoured by R_2 in **Y**. Once again, this response would strike me as plausible.

We can also apply this approach to the case in which our decision maker faces a one-off instance of only the **Special Obligation** choice situation,

followed by neither **X**, nor **Y**, nor even any other possible choice situation. In my preferred IMM model for this instance of **Special Obligation**, the social planner will be initially endowed with the right to determine who the decision maker will be instructed to aid. As always, the social planner's sole objective will be to induce the course of action ϕ that comes close as possible to rendering every theory representative indifferent between (a) repeating ϕ ten times over in response to a sequence of ten identical instances of **Special Obligation**, and (b) splitting the decision rights over this sequence of choice situations 90:10 between R_1 and R_2 .

Of course, only two courses of action are available to the social planner in a one-off instance of **Special Obligation**: (1) instructing our decision maker to aid her parents; or (2) instructing our decision maker to aid the stranger. Clearly, neither of these two courses of action would perfectly satisfy the SEQUENCE CRITERION. However, an instruction to aid the stranger would obviously come much closer to satisfying the SEQUENCE CRITERION than would an instruction to aid the parents. Hence, my preferred version of IMM implies that it is most appropriate for our decision maker to aid the stranger in a one-off instance of **Special Obligation**. After all, this is the course that minimises the shortfall from giving each moral theory its due influence under these circumstances.¹¹

Furthermore, we can also apply this approach to the case in which our

¹¹Some readers might find it helpful to identify the one-off instance of **Special Obligation** as the limiting case of **Special Obligation-then-X** where the amount of money available in **X** is set to zero.

decision maker faces a one-off instance of only the **Trolley** choice situation (introduced at the beginning of the present chapter). In my preferred IMM model for this choice situation, the social planner will be initially endowed with the right to determine whether the decision maker will push the heavysset man, pull the lever, or do nothing. The social planner's sole objective will be to induce the course of action ϕ that minimises the shortfall from rendering every theory representative indifferent between (a) repeating ϕ – say – ten times over in response of a sequence of ten identical instances of **Trolley**, and (b) splitting the decision rights over this sequence of choice situations 50:50 between R_1 and R_2 .

In §5.1 above, I discussed which contract would be agreed to by our theory representatives in a sequence of ten identical instances of **Trolley**. I argued that if the decision rights over the sequence were to be initially divided 50:50 between R_1 and R_2 , and if R_1 and R_2 's bargaining positions were to be structurally identical to each other (in the sense defined in §2.4 above), then R_1 and R_2 would both agree to instruct the decision maker to pull the lever in all ten instances of **Trolley**. Thus, according to the SEQUENCE CRITERION, some course of action ϕ would perfectly satisfy the due influence desideratum in our one-off instance of **Trolley** iff every theory representative would be indifferent between (a) repeating ϕ ten times over in response to a sequence of ten identical instances of **Trolley**, and (b) the decision maker being instructed to pull the lever in all ten instances of **Trolley** in this sequence.

Clearly, instructing our decision maker to pull the lever is the only available course of action that would perfectly satisfy this SEQUENCE CRITERION in a one-off instance of **Trolley**. Hence, as desired (recall §5.4.1 above), my preferred version of IMM implies that it is most appropriate for our decision maker to pull the lever in a one-off instance of **Trolley**. After all, this is the only available course of action that gives each moral theory its due influence under these circumstances.

Thus, in both the **Special Obligation** and **Trolley** one-off discrete choice situations, the social planning version of IMM avoids the implausible implications that I imputed to the lottery version of IMM in §5.3 above. This is because the social planning approach does not introduce an artificial element of risk into the IMM model. Hence, it avoids introducing implausible stochasticities in IMM's appropriateness verdicts, and it does not give risk-averse moral theories less than their due influence. These strike me as decisive advantages of the social planning approach.

5.4.7 Sequences

In §5.1 above, I discussed how IMM should handle sequences of discrete-choice situations that are neatly divisible in proportion to credences. And just now, in §5.4.6 above, I discussed how IMM should handle one-off discrete choice situations. However, I have not yet discussed how IMM should handle sequences of discrete-choice situations that are *not* neatly divisible in proportion to credences.

Fortunately, the social planning approach is also readily applicable to this new class of cases. For instance, suppose our decision maker knows that she is about to confront a sequence of – say – nine identical instances of the **Special Obligation** choice situation. I will call this set of circumstances **SpOb**×**9** for short.

In my preferred IMM model for this **SpOb**×**9** set of circumstances, our social planner will be initially endowed with the right to determine who the decision maker will be instructed to aid in each of these nine instances of **Special Obligation**. And the social planner’s sole objective will be to induce the course of action ϕ in **SpOb**×**9** that minimises the shortfall from rendering every theory representative indifferent between (a) repeating ϕ ten times over – say – in response to a sequence of ten repetitions of **SpOb**×**9**, and (b) splitting the decision rights over this sequence of ten repetitions of **SpOb**×**9** in a 90:10 ratio between R_1 and R_2 . In other words, the planner’s objective will be to select the course of action ϕ in **SpOb**×**9** that minimises the shortfall from rendering every theory representative indifferent between (a) repeating ϕ ten times over in response to a sequence of ten repetitions of **SpOb**×**9**, and (b) splitting the decision rights over a sequence of ninety identical instances of **Special Obligation** in a 90:10 ratio between R_1 and R_2 , as illustrated in figure 5.7.

Under these conditions, it is safe to assume that this SEQUENCE CRITERION would come closest to being satisfied by a course of action in **SpOb**×**9** under which the social planner would instruct the decision maker to aid the

<i>choice situation:</i>	Sp. Obl. 1	...	Sp. Obl. 81	Sp. Obl. 82	...	Sp. Obl. 90
<i>initial assignment of control rights:</i>	R_1	...	R_1	R_2	...	R_2

Figure 5.7: Proportional division in a sequence of ninety instances of **Special Obligation**

stranger in eight out of the nine instances of **Special Obligation**, and would then instruct her to aid her parents in the single remaining instance of **Special Obligation**. Although this course of action would give T_1 slightly less – and T_2 slightly more – than its due influence in **SpOb**×**9**, it nonetheless comes closer than any other available course of action to perfectly satisfying this due influence desideratum. Hence, my preferred version of IMM implies that the most appropriate response to **SpOb**×**9** will involve our decision maker aiding the stranger eight times out of nine, and aiding her parents once. This strikes me as a plausible implication.

I have now demonstrated that my social planning version of IMM is readily applicable to indivisible sequences of discrete-choice situations. However, this social planning approach can in fact also subsume the proportional endowments response to divisible sequence of discrete choice situations that I originally defended in §5.1 above.

For instance, suppose that our decision maker is about to confront a sequence of ten identical instances of **Special Obligation**. I will call this set of circumstances **SpOb**×**10** for short. In my preferred IMM model for

SpOb \times **10**, the social planner will be initially endowed with the right to determine who the decision maker will be instructed to aid in each of these ten instances of **Special Obligation**. And the social planner's sole objective will be to induce the course of action ϕ in **SpOb** \times **10** that minimises the shortfall from rendering every theory representative indifferent between (a) repeating ϕ many times over in response to a repeated sequence of instances of **SpOb** \times **10**, and (b) splitting the decision rights over this sequence of choice situations 90:10 between R_1 and R_2 .

Clearly, however, our social planner can perfectly satisfy this SEQUENCE CRITERION in **SpOb** \times **10** simply by initially endowing R_1 with the right to decide how the decision maker will behave in nine instances of **Special Obligation** from this set of ten, with R_2 being endowed with this right in the sole remaining instance of **Special Obligation**. As desired, then, my social planning proposal has the further advantage of underwriting the most natural IMM response to perfectly divisible sequences of discrete choice situations.

5.5 Complications

5.5.1 Repetition

Throughout this chapter, I have been implicitly assuming that for any given set of circumstances Φ , there must exist some logically possible sequence of choice situations wherein Φ is repeated many times over. Furthermore, I have

also explicitly assumed that according to every moral theory in which our decision maker has positive credence, the choiceworthinesses of the options available in any given choice situation do not depend upon the sequence of choice situations that this decision maker has previously confronted, or upon how the decision maker has already behaved in any of these choice situations. Call these my ‘repeatability assumptions.’

My development of the SEQUENCE CRITERION for due influence relied upon both of these repeatability assumptions. For one thing, this CRITERION is intelligible only under the first of these two assumptions. Moreover, in my argument to motivate the SEQUENCE CRITERION, I defended the PROPORTIONALITY CLAIM by means of an analogy between (1) the control rights over the discrete choice situations in a repeated sequence, and (2) the control rights over individual resource units in a division problem like **Philanthropy** with constant returns to scale. And this analogy makes sense only under the assumption that the choiceworthinesses of the options available in any given discrete choice situation do not depend upon how our decision maker has already behaved in previous choice situations.

Unfortunately, both of these two repeatability assumptions appear at first glance appear quite implausible. Firstly, there are some moral choice situations for which there would not seem to exist any logically possible sequence of repetitions. For example, imagine that our decision maker has to decide whether or not to make the duck-billed platypus permanently extinct. Clearly, this is not a choice situation that one could imagine repeating multi-

ple times over, since no species can be rendered permanently extinct multiple times.

Secondly, our decision maker might also have positive credence in some theories according to which the choiceworthinesses of the options available in any given discrete choice situation depend upon how our decision maker has already behaved in previous choice situations. For instance, imagine that our decision maker has credence in some moral theory according to which the choiceworthiness of choosing to kill in **Trolley** depends upon whether this decision maker has ever killed anyone before. Perhaps according to this moral theory, one's first act of killing is more unchoiceworthy than any subsequent acts of killing, because one's first act of killing permanently stains one's moral character. But then of course according to this moral theory, the choiceworthinesses of pulling the lever or pushing the heavysset man in any given instance of **Trolley** will depend upon how the decision maker behaved in any previous instances of that choice situation.

Fortunately, however, we can rescue both of my repeatability assumptions from these objections by reconceptualising the notion of a 'choice situation' in far more abstract terms than one might at first have expected. For instance, imagine that our decision maker faces a choice between (i) pushing a heavysset man into the path of a trolley, (ii) pulling a lever to redirect the trolley towards two people, and (iii) doing nothing, allowing five to die. Suppose that according to T_1 , pushing the heavysset man has a choiceworthiness value of $+1 T_1$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9$

T_1 -units, and doing nothing has a negative choiceworthiness value of $-1 T_1$ -units. Furthermore, according to T_2 , doing nothing has a choiceworthiness value of $+1 T_2$ -units, whereas pulling the lever has a choiceworthiness value of $+0.9 T_2$ -units, and pushing the heavysset man has a negative choiceworthiness value of $-1 T_2$ -units.

I want to suggest that the abstract facts about choiceworthiness values according to T_1 and T_2 are the only facts that we should understand as being essential to the ‘choice situation’ confronted by this decision maker. In other words, I want to suggest that the following description captures everything that we should regard as essential to the decision maker’s choice situation:

- (1) There are three options available.
- (2) The first of these three options has a choiceworthiness value of $+1 T_1$ -units according to T_1 , but of $-1 T_2$ -units according to T_2 .
- (3) By contrast, the second option has a choiceworthiness value of $+0.9 T_1$ -units according to T_1 , and of $+0.9 T_2$ -units according to T_2 .
- (4) Finally, the third of these three options has a choiceworthiness value of $-1 T_1$ -units according to T_1 , but of $+1 T_2$ -units according to T_2 .

Thus, a world in which our decision maker’s current ‘choice situation’ is repeated N times over would simply be a world in which our decision maker

faces N-many choice situations in which there are three options available, and the choiceworthinesses of these three options are described by the propositions (2) through (4). So there is really no need to get bogged down in actually imagining whether any of these three options is the option to ‘kill,’ or to ‘kill for the first time,’ or to ‘make the duck-billed platypus extinct’ – or indeed anything like that.

According to this new reconceptualisation, ‘choice situations’ are individuated exactly and only by the choiceworthinesses of the options available in them according to all of the moral theories in which our decision maker has positive credence. Thus, if our decision maker has positive credence in some moral theory according to which killing in the first occurrence of **Trolley** that she faces would be morally worse than killing in a second occurrence of **Trolley**, then my new approach to individuating choice situations implied that in this second occurrence of **Trolley**, our decision maker does *not* in fact face an instance of ‘the same choice situation’ as she did in our first occurrence of **Trolley**. Fortunately, however, this result in no way prevents us from designing a logically possible sequence of choice situations in which the particular choice situation faced by our decision maker in the first occurrence of **Trolley** is repeated many times over. After all, all that we need to do to design this sequence is to repeat many times over the abstract choice structure which is instantiated by the first occurrence of **Trolley**.

My reconceptualisation of the notion of a ‘choice situation’ also abstracts away from any particular features that might threaten the logical possibility

of any given sequence of choice situations. For example, although it is logically impossible to render the duck-billed platypus permanently twice over, there could be nothing logically impossible about making two choices in sequence whose choiceworthinesses according to every moral theory in which our decision maker has positive credence would both be equal to choiceworthiness of rendering the platypus permanently extinct.

Hence, the SEQUENCE CRITERION for due influence need not be threatened by any concerns about purportedly ‘unrepeatable’ choice situations.

5.5.2 Division

Thus far in this chapter, I have also been implicitly assuming that our decision maker’s credence in any given moral theory must be a rational number – such as $\frac{1}{10}$, $\frac{1}{2}$ or $\frac{9}{10}$. However, some decision makers might instead have credences in certain moral theories that are irrational numbers – like $\frac{1}{\pi}$, or $\sqrt{0.5}$.¹²

This possibility of irrational-number credences creates a new problem for the applicability of my SEQUENCE CRITERION for due influence. Recall that according to the SEQUENCE CRITERION,

any given course of action ϕ in response to any given set of circumstances Φ would perfectly satisfy the due influence desideratum IFF every theory representative would be indifferent be-

¹²A more general problem for IMM concerns *continuous* moral credence distributions. I will discuss this more general problem in my future work.

tween (a) repeating ϕ many times over in response to a repeated sequence of instances of Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories.

However, if our decision maker has at least some irrational number credences, then it will be impossible to split any finite sequence of discrete choice situations between her theory representatives exactly in proportion to the decision maker's credences. Hence, under these circumstances the social planner for our decision maker cannot use the SEQUENCE CRITERION to determine which courses of action would count as giving each moral theory its due influence.

Fortunately, however, we can avoid this problem by making a natural and well-motivated fix to the SEQUENCE CRITERION. This fix is motivated by the following revised version of my PROPORTIONALITY CLAIM from §5.4.2 above:

PROPORTIONALITY*: in a world in which any given set of circumstances Φ is repeated N -many times, splitting the decision rights over this sequence of choice situations between theory representatives in a way that comes as close as possible to matching the decision maker's credences in the corresponding moral theories will, as N increases, tend towards inducing a course of action that perfectly satisfies the due influence desideratum.

After all, as N tends towards infinity, the distribution of decision rights that comes as close as possible to matching the decision maker's credence distribution over moral theories will tend towards identity with that credence distribution.

Taking PROPORTIONALITY* together with the REPETITION and INDIFFERENCE CLAIMS from §5.4.2 above, we can deduce the following revised version of the SEQUENCE CRITERION for due influence:

SEQUENCE*: any given course of action ϕ in response to any given set of circumstances Φ would perfectly satisfy the due influence desideratum IFF as N increases, every theory representative will tend towards being indifferent between (a) repeating ϕ many times over in response to a repeated sequence of N -many instances of Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in a way that comes as close as possible to matching the decision maker's credences in the corresponding moral theories.

And with this limited fix onboard, SEQUENCE* is applicable even for decision makers who have irrational-number credences in some moral theories.

5.6 Synthesis

In this chapter, I have argued that in order to handle discrete-choice situations, IMM model should be augmented by an imaginary social planner agent,

whose sole objective is to ensure that every theory representative comes as close as possible to having its due influence on the decision maker's overall life plan. I have argued that this social planning version of IMM has plausible implications in a wide range of cases, and compares favorably to the lottery alternative.

Introducing the social planner is also an entirely natural extension of my original motivation by analogy for the IMM response to moral uncertainty. In fact, the social planning version of IMM merely explicitly introduces into the IMM model one element that I had until now left implicit. After all, I noted in §1.3 above that a property rights plus markets based response to the *social* problem of incompatible *desires* becomes more attractive if we can imagine that the initial distribution of property is “completely determined by some social planner whose social objective is to give each household its ‘fair share’ of property rights.” Analogously, in the IMM response to moral uncertainty, the initial distribution of control rights should be completely determined by some social planner whose social objective is to ensure that every theory representative comes as close as possible to having its due influence on the decision maker's overall life plan.

In resource division choice situations like **Philanthropy**, I could leave this aspect of the IMM model implicit, instead simply stipulating that control rights over resources should be divided between the theory representatives in proportion to the decision maker's credences in the corresponding moral theories. By contrast, however, in discrete-choice situations like **Trolley**, the

social planner needs to be introduced into the IMM model explicitly, since in discrete choice situations there will be no scheme for initially endowing theory representatives with decision rights which will satisfy the desideratum of due influence. Hence, the social planner should herself keep control of these decision rights in discrete-choice situations.

Of course, even in this social planning version of the IMM approach, appropriateness judgments are still ultimately determined by models in which control rights are initially divided among theory representatives in proportion to credences, and then after that allowing the representatives to negotiate trades and contracts with each other. After all, the SEQUENCE CRITERION for due influence is defined in terms of the course of action that would be induced by splitting the decision rights over some sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories, and then after that allowing the representatives to negotiate trades and contracts with each other.

In fact, my original proportional endowments response to resource division choice situations can now be entirely subsumed by the social planning approach. Let me now stipulate that whenever our decision maker encounters *any* set of circumstances Φ , our social planner will be initially vested with control over the decision maker's behaviour in that set of circumstances. The social planner can then decide either to exercise these control rights herself, or otherwise to distribute those control rights amongst the theory representatives. As always, the social planner's sole objective will be to induce the

course of action ϕ that comes as possible to rendering every theory representative indifferent between (a) repeating ϕ many times over in response to a repeated sequence of instances of Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories.¹³

Suppose, for instance, that our decision maker simply confronts a one-off instance of the **Philanthropy** choice situation (introduced in §1.3 above). Then in my preferred IMM model for this set of circumstances, the social planner will be initially endowed with control rights over our decision maker's entire fortune in the **Philanthropy** choice situation. Furthermore, the social planner's social objective will be to induce the course of action ϕ that comes as close as possible to rendering every theory representative indifferent between (a) repeating ϕ many times over in response to a repeated sequence of instances of **Philanthropy**, and (b) splitting control over the total stock of resources in this sequence of choice situations 60:40 between R_1 and R_2 .

Notice, however, that one way to ensure a 60:40 division between R_1 and R_2 in control over the total stock of resources in a repeated sequence of instances of **Philanthropy** would simply involve imposing a 60:40 division between R_1 and R_2 in control rights over the resource endowment in each of the individual **Philanthropy** choice situations. Hence, our social planner

¹³Strictly speaking, this statement of the SEQUENCE CRITERION should be replaced by the revised SEQUENCE* criterion that I suggested in §5.5.2 above. However, for ease of exposition, from now on I will always just use the original version of SEQUENCE CRITERION. This is a harmless simplification.

can perfectly satisfy the SEQUENCE CRITERION in a one-off instance of **Philanthropy** simply by initially granting a 60:40 division between R_1 and R_2 in control rights over the resource endowment in that choice situation. This is just one instance of the more general result that our social planner will underwrite proportional endowments of resources to the theory representatives in resource division choice situations like **Philanthropy**.

In light of this result that the social planning approach can subsume the proportional endowments response to resource division choice situations, we can now give a concise (re)statement of my preferred IMM theory of appropriateness. First of all, consider the special case in which our decision maker is certain that Φ is the only set of circumstances she will ever encounter wherein any of her theory representatives issue incompatible directives. Recall that granted this assumption, we can ignore the complexities introduced by intertemporal bargaining (discussed in §4 above). Furthermore, let us also assume for the sake of simplicity that our decision maker's credence in any given moral theory must be a rational number. (Granted this assumption, we can ignore the gratuitous complexities introduced by irrational-number credences, discussed in §5.5.2 above.)

Granted these assumptions, IMM says that

the most appropriate course of action in response to Φ will be the course of action ϕ that comes as close as possible to rendering every theory representative indifferent between (a) repeating ϕ many times over in response to a repeated sequence of instances of

Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's credences in the corresponding moral theories.

More generally – to take into account issues of intertemporal dynamics (recall chapter 4 above) – my preferred version of IMM says that

IF

- Ψ is the set of choice situations that our decision maker has confronted thus far in her lifetime;
- ψ is the set of choices that she made in those situations;
- and Φ is the set of circumstances that our decision maker faces in deciding which plan to follow for the remainder of her lifetime,

THEN the most appropriate course of action in response to Φ will be the course of action ϕ that comes as close as possible to rendering every theory representative indifferent (given their current descriptive credences) between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker's (current) credences in the corresponding moral theories.

My primary aim in this chapter has been to develop an attractive IMM response to discrete choice situations. Yet in so doing, I also hope to have bolstered the overall appeal of my broader IMM project. I have argued in this section that the best IMM response to discrete choice situations turns out to cohere elegantly with the best IMM response to resource division scenarios. This goes to show my IMM emerging approach evinces the sort of theoretical virtues that we should want it to. Far from being an *ad hoc* modification to the IMM framework, my social planning response to discrete choice situations is fully grounded in IMM's core ideas.

Chapter 6

Prerogatives

In every choice situation that I have considered thus far in this dissertation, it has been easy to work out which preferences each theory representative should have if it is certain in the truth of the moral theory that it represents, and wants our morally uncertain decision maker to follow the directives of that particular moral theory as closely as possible (as stipulated in §1.3). For instance, in the **Philanthropy** choice situation, T_1 implied that our philanthropist should donate as much as possible to deworming, whereas T_2 implied that she should donate as much as possible to soup kitchens. Thus, we could straightforwardly infer that R_1 and R_2 prefer for the philanthropist to donate as much as possible to, respectively, deworming and soup kitchens.

Yet it might be less clear what preferences an IMM theory representative should have if it is to represent some moral theory which incorporates *agent-centred prerogatives*. For instance, consider the following choice situation:

Operations: some decision maker is deciding how to divide her life savings. This decision maker must distribute her savings between two possible uses: the first of which is helping to fund an operation for a sick close friend of hers, and the second of which is helping fund similar operations for several distant strangers. Suppose that this decision maker has 99% credence in some moral theory T_1 , and 1% in another moral theory T_2 . According to the theory T_1 , moral agents have a strong agent-centred prerogative over how they use their own savings. Hence, T_1 implies that any possible distribution of money between aiding the friend and aiding the strangers would be morally permissible under these conditions. By contrast, according to T_2 , our decision maker is morally required to use all of her life savings to fund operations for the distant strangers under these circumstances.

Let us assume for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to **Operations**. Moreover, let us also assume that the choiceworthiness returns to scale from aiding the friend and aiding the stranger are *constant* according to both of the moral theories T_1 and T_2 .

It might at first seem natural to suppose that T_1 's theory representative R_1 should be indifferent over how our decision maker distributes her savings

in **Operations**.¹ After all, T_1 implies that every available option is permissible. Hence, it seems as though T_1 's directives do not give R_1 a reason to prefer or disprefer any particular distribution of her savings.

As it happens, however, I will argue that we should ultimately reject this apparently plausible assumption that prerogatives can be modelled as sources of indifference.² To see why this is so, suppose for the sake of *reductio* that R_2 should be modelled as being indifferent about how our decision maker in **Operations** distributes her savings. Furthermore, suppose for the sake of simplicity that our decision maker is certain to never encounter any choice situations other than **Operations** in which T_1 and T_2 disagree, thus ruling out the potential for any intertemporal trades or contracts. Under these conditions, donating everything to the distant strangers would weakly Pareto dominate every other distribution, with respect to R_1 and R_2 's preferences over this moral choice. After all, donating everything to the strangers is no worse than any other possible distribution according to R_1 , yet it is the best possible distribution according to R_2 . Hence, any version of IMM whose bargaining solution rules out weakly Pareto dominated outcomes must imply that donating everything to the distant strangers is the only appropriate option under these conditions.

This implication strikes me as highly unattractive. Our decision maker has 99% credence in the existence of a strong agent-centred prerogative guar-

¹Kaczmarek, Lloyd and Plant 2025, §4.5.

²*Pace* Kaczmarek, Lloyd and Plant 2025, §4.5.

anteeing that it would be morally permissible for her to donate some or all of her savings to her close friend. In light of this fact, any plausible version of IMM should imply that this decision maker also has a strong ‘second-order’ prerogative guaranteeing that it would be *appropriate* for her to donate at least some of her savings to her close friend.

Thus far, I have demonstrated that if R_1 is stipulated to be indifferent between all of the available distributions in **Operations**, then any version of IMM whose bargaining solution rules out weakly Pareto dominated outcomes will imply that donating everything to the distant strangers is the only appropriate option in this choice situation. Hence, one might now wonder whether we could avoid this implausible implication by instead adopting a version of IMM whose bargaining solution does *not* rule out weakly Pareto dominated outcomes.³ Unfortunately, however, I will now argue that all of these versions of IMM also have several of their own implausible implications.

Firstly, these versions of IMM have implausible implications in choice situations where one or more of our theory representatives is indifferent between two available options for reasons having nothing to do with an agent-centred prerogatives. For instance, consider the following choice situation:

Dominance: some decision maker is choosing between only two possible options, A and B. Suppose that this decision maker has positive credence in only two different moral theories – one of which is total-utilitarian, and the other is deontological. As it

³Kaczmarek, Lloyd and Plant 2025, §4.5.

happens, options A and B would both produce exactly the same amount of total wellbeing, and so the utilitarian theory implies that these two options are equally choiceworthy. By contrast, however, the deontological theory implies that A is the only permissible option in this choice situation.

Any version of IMM whose bargaining solution does not rule out weakly Pareto dominated outcomes in **Dominance** will imply that options A and B are *both* appropriate for this decision maker. On the contrary, however, it strikes me as highly plausible to suppose that A is the only appropriate option available in **Dominance**. In other words, this is a choice situation in which we *should* want our IMM bargaining solution to rule out the weakly dominated option B.

Furthermore, suppose we continue to stipulate (for *reductio*) that R_1 is indifferent between all of the available distributions in **Operations**. Then under these conditions, even versions of IMM whose bargaining solutions do not rule out weakly Pareto dominated outcomes can still have implausible implications in some sets of circumstances that involve **Operations**. For instance, suppose our decision maker knows that after she has decided what to do in **Operations**, she will then be confronted by an instance of the following choice situation:

Free Time: our decision maker has one hour of free time this afternoon, and now faces a choice between only two uses for it:
 (1) visiting her parents for an hour; or (2) spending an hour

volunteering at some local charity. Suppose that according to T_1 , our decision maker should spend her free time visiting her parents; whereas according to T_2 , she should spend her time volunteering at the local charity.

Finally, suppose for the sake of simplicity that our decision maker is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade of contracts involving other scenarios in addition to **Operations** and **Free Time**.

In **Free Time**, R_1 will prefer for our decision maker to visit her parents, whereas R_2 will prefer for her to volunteer at the local charity. However, we can safely assume that the stakes in **Free Time** are much lower than they are in **Operations** according to R_2 . Hence, if we stipulate (for *reductio*) that R_1 is indifferent between the two options available in **Operations**, then we can safely infer that in **Operations-then-Free Time**, the decision maker

(a) donating all of her savings to her close friend, and then spending her free hour volunteering at the local charity

will be *strictly* Pareto dominated by the alternative course of action

(b) donating all of her savings to the distant strangers, and then spending her free hour visiting her parents.

After all, R_1 only cares about how our decision maker chooses to spend her free time, whereas R_2 cares more strongly about how our decision maker chooses to use her savings.

Under these conditions, no plausible version of IMM could imply that it is appropriate in **Operations-then-Free Time** for our decision maker to donate all of her savings to her close friend and then spend her free hour volunteering at the local charity. After all, no set of rational bargainers would ever settle for a strictly Pareto-dominated outcome. Hence, no plausible specification of IMM's bargaining process could select an option that is strictly Pareto dominated with respect to our theory representatives' preferences.

Unfortunately, however, donating to the friend and then volunteering at the local charity does in fact strike me as one potentially appropriate response to the **Operations-then-Free Time** set of circumstances. Our decision maker has 99% credence in a partialist moral theory T_1 according to which moral agents have a strong agent-centred prerogative over how they use their own savings, and only 1% in the impartialist moral theory T_2 . Hence, donating to the friend but then volunteering at the local charity strikes me as a response to **Operations-then-Free Time** which could potentially afford due influence (in proportion to credences) to both of these moral theories. Any version of IMM which implies that this course of action must be inappropriate would *ipso facto* be giving T_1 less than its due influence on our decision maker's overall lifetime course of behaviour.

This concludes my argument against the idea that T_1 's theory representative R_1 should be modelled as being indifferent over all of available donation distributions in **Operations**. Having now argued against this *prima facie* natural suggestion, I will spend the remainder of this section developing and

defending an alternative IMM approach to agent-centred prerogatives.

The basic idea behind this alternative approach to agent-centred prerogatives is that we should think about **Operations** as a choice situation in which R_1 's preferences are *underdetermined* by T_1 's directives to our decision maker. In other words, we should say that wanting our morally uncertain decision maker to follow T_1 's directives is in fact consistent with *any* set of preferences over of the available donation distributions in **Operations**. T_1 's directives are simply not restrictive enough to determine the preferences that R_1 should have in **Operations**.

Suppose we stipulate that R_1 's preferences are underdetermined for our IMM bargaining model of any given set of circumstances involving **Operations**. Then under these conditions, it will likely also be underdetermined what the theory representatives would decide to do with their endowments of control rights in this IMM model, and which (if any) trades or contracts they would agree to. For instance, suppose once again for the sake of simplicity that our decision maker is certain to never encounter any choice situations in which T_1 and T_2 disagree, other than one single instance of **Operations**. Then under these conditions, IMM implies that appropriateness of any given donation distribution should depend upon whether our decision maker could be instructed to select this donation distribution by the theory representatives R_1 and R_2 in an IMM model of **Operations** where R_1 is initially endowed with control rights over 99% of the decision maker's savings, with R_2 being initially endowed with control rights over the remaining 1%.

But of course, the instructions issued to our decision maker in this IMM model will depend upon R_1 's preferences over the options in **Operations**. For instance, if R_1 preferred for our decision maker to donate as much as possible to her close friend, then R_1 would instruct our decision maker to donate 99% of her savings to the close friend, with R_2 instructing her to donate the remaining 1% of her savings to distant strangers. However, if R_1 instead preferred for our decision maker to donate as much as possible to distant strangers, then R_1 and R_2 would together instruct our decision maker to donate 100% of her savings to the strangers. Finally, if R_1 were to be indifferent over all of the donation distributions available in **Operations**, then the instructions issued to our decision maker will depend upon exactly how we model the IMM inter-representative bargaining process (a topic that I will discuss in §7 of this dissertation).

Hence, if R_1 's preferences are underdetermined in **Operations**, then the instructions issued to our decision maker by R_1 in the corresponding IMM decision model will also be underdetermined. Hence, in order to handle this set of circumstances, we now need to stipulate what IMM's appropriateness verdicts should be in cases like this where it is underdetermined what course of action IMM's theory representatives would jointly select.

One possible option might be to stipulate that in any cases like this with underdetermined model instructions, IMM should simply imply that it is also in some sense '*indeterminate*' which courses of action would be appropriate for our decision maker. For instance, in **Operations**, donating everything to

the distant strangers would be neither determinately appropriate, nor determinately inappropriate. Likewise, a 99:1 split split between the close friend and the distant strangers would also be neither determinately appropriate, nor determinately inappropriate.

In my view, however, this possible stipulation would be unattractive. It strikes me as *prima facie* plausible to suppose that donating everything to the distant strangers, for instance, must be a *determinately* appropriate option in **Operations**. After all, our decision maker has 99% credence in a moral theory T_1 according to which donating everything to the stranger is permissible, with her remaining 1% credence being in the moral theory T_2 according to which this option is required in **Operations**. Thus, it is difficult to see how donating everything to the stranger could be anything less than determinately appropriate. (In fact, it is not even clear to me what it could mean to say that some course of action is indeterminate in its appropriateness.)

In order to honour these intuitions, I suggest that we should adopt a *subvaluationist* account of appropriateness in choice situations like this wherein some theory representatives' preferences are underdetermined by the directives of the corresponding moral theories.⁴ First of all, let us say that any particular specification of complete and determinate preferences for any given

⁴I call my account 'subvaluationist' because it will be *loosely* structurally analogous to subvaluationist theories of vagueness (cf. Cobreros 2013). Leora Sung (2023) has also proposed another (rather different) structurally subvaluationist account of appropriate choice under uncertainty over moral theories which incorporate some prerogatives.

IMM theory representative can be referred to as a ‘*preference assignment*.’ Furthermore, let us also say that a preference assignment would count as ‘*legitimate*’ for any given theory representative iff this representative having these assigned preferences would be compatible with her wishing for our decision maker to follow the directives of the moral theory corresponding to this representative. Now, according to my preferred subvaluationist theory of appropriateness, any given donation distribution ϕ in response to **Operations** is maximally appropriate iff there exists *at least one* set of legitimate preference assignments under which our IMM theory representatives would jointly instruct our decision maker to choose ϕ .⁵

According to this subvaluationist account of appropriateness, a donation distribution is inappropriate in **Operations** iff this distribution allocates less than 1% of the decision maker’s savings to aiding distant strangers. After all,

⁵More generally, my preferred subvaluationist version of IMM says that (recall §5.6 above)

IF

- Ψ is the set of choice situations that our decision maker has confronted thus far in her lifetime;
- ψ is the set of choices that she has made in those situations;
- and Φ is the set of circumstances that our decision maker faces in deciding which plan to follow for the remainder of her lifetime

THEN any given course of action ϕ is a maximally appropriate response to Φ iff there exists at least one set of legitimate IMM preference assignments under which ϕ minimises the shortfall from rendering every theory representative at worst indifferent (given their current descriptive credences) between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) splitting the decision rights over this sequence of choice situations between theory representatives in proportion to the decision maker’s (current) credences in the corresponding moral theories.

regardless of how we specify R_1 's preferences over donation distributions, we can always be sure that R_2 will use her initial endowment of control rights over 1% of our decision maker savings to instruct her that this 1% be donated to the distant strangers. But on the flipside, how R_1 will want to use her initial endowment of control rights over 99% of the decision maker's savings is left undetermined by T_1 's directives, and so any way of allocating this 99% can count as appropriate according to my subvaluationist account.

Overall, this strikes me as an attractive approach to handling agent-centred prerogatives. In **Operations**, our decision maker has 99% credence in the existence of a strong agent-centred prerogative guaranteeing that it would be morally permissible for her to donate some or all of her savings to her close friend. Hence, under these conditions, my subvaluationist version of IMM implies that this decision maker thus has a strong 'second-order' prerogative guaranteeing that it would be appropriate for her to donate anything up to 99% of her life savings to her close friend. Both in this case and in general, this strikes me as a desirable approach to handling first-order prerogatives.

Chapter 7

Bargaining

7.1 Cooperative solutions

At several points in this dissertation thus far, I have mentioned that IMM's exact implications for appropriate choice in some situations (such as **Three Charities**, **Double Distribution**, and **Trolley**) can depend upon exactly how we model the details of the bargaining process between theory representatives. In this chapter, I will now turn my attention to this theoretical choice point. My aim here will not be to conclusively settle this particular question. So, I won't try to prove that the bargaining model which I defend in this chapter is necessarily the best bargaining model for use with IMM. Rather, my more modest aim will just be to show that there is at least one bargaining model under which IMM is more attractive than rival theories of appropriateness like MEC, and hence to show that IMM represents a signif-

icant theoretical step in the right direction. This chapter will necessarily be somewhat technical, and so some readers may simply prefer to skip ahead now to the next chapter.

7.1.1 Nash bargaining

In certain choice situations – such as, for instance, **Three Charities** and **Double Distribution** – there could be multiple possible Pareto-optimal contracts that would each be preferable over the *status quo* (no-contract) outcome, according to every one of our IMM theory representatives (recall §§2.4-2.5 above). Hence, in order to handle these kinds of circumstances, IMM needs to incorporate a precise bargaining ‘solution concept’ that can determine which contracts our theory representatives would agree to.

Perhaps the most well-known such solution concept in formal bargaining theory is the *Nash bargaining solution* (henceforth: ‘NBS’). In his groundbreaking treatment of the bargaining problem, John Nash laid out four plausible axioms on the final outcomes of good-faith (referred to as ‘co-operative’) bargaining procedures:¹

1. *Scale invariance*: any positive affine rescaling of any bargainers’ utility functions should not alter the bargaining solution (I will discuss and explain this axiom in §§7.1.3-7.1.5 below).
2. *Pareto optimality*: no feasible alternatives should Pareto dominate the

¹Nash 1950.

bargaining solution

3. *Symmetry*: if all bargainers have the same utility functions, and would all have the same *status quo* utilities if bargaining broke down, then they should all have identical utilities under the bargaining solution
4. *Independence of irrelevant alternatives*: eliminating any given set of possible outcomes should only make a difference to the bargaining solution if one of the eliminated outcomes would itself have been selected as the bargaining solution had it not been eliminated

The NBS uniquely satisfies all four of these axioms.² But before I can state the NBS, I will first of all need to introduce some new notation. Firstly, let $\{1, \dots, n\}$ denote the set of bargainers. Then for each bargainer $i \in \{1, \dots, n\}$, let $u_i(a)$ denote i 's utility under the possible bargaining outcome a , and let d_i denote i 's *status quo* or 'disagreement' utility.³

Now, some outcome A is a Nash bargaining solution of any given bargaining problem iff setting $a = A$ maximises the Nash maximand

$$\prod_{i=1}^n (u_i(a) - d_i)$$

subject to the constraint that $u_i(a) \geq d_i$ for every bargainer $i \in \{1, \dots, n\}$.⁴

One attractive feature of the NBS is that (all else being equal) it favours

²Nash 1950.

³I will discuss these disagreement utilities in §7.1.2.

⁴For those unfamiliar with this notation, ' $\prod_{i=1}^n \theta_i$ ' abbreviates ' $\theta_1 \times \theta_2 \times \dots \times \theta_n$ ' (for any given $\theta_1, \theta_2, \dots, \theta_n$).

equal divisions of the gains from trade between bargainers. For example, suppose that two bargainers are choosing between (i) an option A that gives each bargainer a utility gain of 4 over the *status quo*, and (ii) an alternative option B that gives the two bargainers utility gains of 2 and 6 respectively. For option A, the value of the Nash maximand is $4 \times 4 = 16$, whereas for option B the value of the Nash maximand is $2 \times 6 = 12$. Hence, as desired, the NBS prefers option A over option B.

7.1.2 Disagreement utilities

Formal bargaining solutions like the NBS must always be calculated relative to a set of *status quo* or ‘disagreement’ utilities d_1, \dots, d_n which each of our bargainers would have if the bargaining process broke down without the bargainers reaching any agreements. Thus, if we wish to apply any of these bargaining solutions to our IMM model of any given sequence of choice situations, then we will first of all have to specify how these disagreement utilities are to be determined.

Roughly speaking, my general idea here is that each theory representative’s disagreement utility in the IMM bargaining model for any given sequence of choice situations should be the utility that this representative would end up with if none of our representatives agreed to any trades or contracts with each other in this IMM model. For example, recall that in the **Three Charities** choice situation (first introduced in §2.4 above), some philanthropist has 50% credence in each of the two moral theories T_1 and T_2 ,

and needs to decide how to distribute her fortune between the three charities A, B, and C. Now imagine that R_1 and R_2 do not agree to any trades or contracts with each other in the IMM bargaining model for **Three Charities**. Under these conditions, recall from §2.4 above that it will be in R_1 's best interests to instruct our philanthropist to donate R_1 's half of her fortune to charity A; and it will be in R_2 's best interests to instruct our philanthropist to donate R_2 's half of her fortune to charity C. Hence, the disagreement utilities in **Three Charities** should be the utilities which R_1 and R_2 would have if our philanthropist split her fortune 50:50 between the charities A and C.

In this particular choice situation, only one set of possible instructions (*viz.* the 50:50 split between A and C) could be rationally issued by our theory representatives if none of these representatives agreed to any trades or contracts. Thus, this is a choice situation in which we can quite straightforwardly apply the basic idea that each theory representative's disagreement utility in any given set of circumstances should just be the utility that this representative would have in our IMM bargaining model if there were no trades or contracts.

Unfortunately, however, we can also imagine some other sequences of choice situations for which there would be multiple difference sets of possible instructions that would each be rationally issued by our set of theory representatives if none of these representatives had agreed to any trades or contracts. For instance, consider any sequence of choice situations in which one or more of our theory representatives would be *indifferent* between mul-

tiple possible uses for their initial endowments of control rights if none of our representatives had agreed to any trades or contracts. In a sequence of choice situations like this, the theory representatives in this position could quite rationally instruct our decision maker to use those representatives' initial endowments of control rights for any of the purposes over which those representatives are indifferent. Hence, under these conditions there would no single determinate answer to the question of which particular set of instructions our theory representatives would jointly issue to our decision maker if we assumed that they could not agree to any trades or contracts with each other.

In *some* – but not all – possible sequences of choice situations like this, every one of our theory representatives might nonetheless be indifferent between all of the possible sets of instructions which could each be rationally issued to our decision maker. Hence, even in some sequence of choice situations like this, there might still be some single determinate answer to the question of which particular set of *utilities* our theory representatives would have if they could not agree to any trades or contracts with each other. Thus, in *these* possible sequences of choice situations, we could still straightforwardly apply the basic idea that our theory representative's disagreement utilities should just be the utilities that they would have if they could not agree to any trades or contracts.

Unfortunately, however, there are also some other sequences of choice situations in which at least some of our theory representatives would *not* be

indifferent between all of the possible sets of instructions which could each be rationally issued to our decision maker. For example, imagine a sequence of choice situations in which R_1 but *not* R_2 or R_3 is indifferent between multiple possible uses for R_1 's endowment of control rights. Under these conditions, there cannot be any single determinate answer to the question of which particular set of utilities our theory representatives would have if they could not agree to any trades or contracts with each other. Hence, in any one of these sequences of choice situations, we cannot straightforwardly apply my basic idea that each theory representatives' disagreement utility should just be the utility that this representative would have if there were no trades or contracts. Thus, I will need to redefine this basic idea in order to handle these sorts of choice situation sequences.

One possible line of response to this challenge would involve stipulating that each theory representative's disagreement utility in any given sequence of choice situations should be calculated as some function of all of the different utility values that this particular representative could have in the version of our IMM bargaining model for this sequence of choice situations in which there are no trades or contracts. More specifically, each theory representatives' disagreement utility could perhaps be calculated by taking an *average* over all of her possible 'no contract' utility values. Or, perhaps we could instead take the *minimum* over these values, or else some other more complicated calculation. Any version of this potential line of response would always ensure that each of our theory representatives is assigned exactly one

disagreement utility value, even for sequences of choice situations in which some of these representatives could have multiple potential ‘no contract’ utilities.

However, a second possible contrasting response to the challenge posed by these sorts of choice situation sequences actually would *not* always involve assigning to each of our theory representatives exactly one disagreement utility. Rather, this alternative line of response is inspired by my *subvaluationist* response to the problem of *prerogatives* (developed in §6 above). According to this new subvaluationist response to the problem of underdetermined disagreement utilities, any given course of action ϕ should be counted as one possible NBS of our IMM bargaining model for any given sequence of choice situations Φ iff there exists *at least one* possible ‘no contract’ profile of utilities $\langle d_1, \dots, d_n \rangle$ under which setting $a = \phi$ would maximise the Nash maximand $\prod_{i=1}^n (u_i(a) - d_i)$.

It is far from obvious to me which of these two possible lines of response to the problem of undetermined disagreement utilities ought to be incorporated into IMM. Hence, I will not attempt to adjudicate this particular theoretical choice point here. Fortunately, none of my claims about IMM in the remainder of this dissertation will depend upon how we settle this choice point, since I will only ever consider sequences of choice situations for which there will always one single determinate answer to the question of exactly which particular sets of utilities our theory representatives would end up with if they could not agree to any trades or contracts with each

other. This means that henceforth I will always be able to straightforwardly apply my basic idea that each theory representative's disagreement utility in the IMM bargaining model for any given sequence of choice situations should just be the utility that this representative would end up with if none of our representatives could agree to any trades or contracts with each other in this IMM model.

7.1.3 Scale invariance

Perhaps I should now explain the significance of Nash's scale invariance axiom (introduced in §7.1.1 above). Let me begin by reviewing the distinction between '*interval-scale*' and '*merely ordinal*' utility functions.

Speaking generally, an '*interval scale*' is any scale such that any given unit increase in the value of this scale always represents an increase of a certain fixed amount in the thing being measured, regardless of whereabouts on that scale the unit increase occurs. For instance, °F and °C are both interval-scale measures of temperature, since an increase of 1°F or 1°C always each represents a certain fixed amount of extra heat, regardless of whether this unit increase occurs at freezing point, boiling point, or any other temperature.

On the other hand, a '*merely ordinal*' scale is any scale for which any given unit increase in the value of the scale does *not* always represent an increase of a certain fixed amount being measured regardless of whereabouts on that the unit increase occurs. For instance, an undergraduate's exam rank relative to her peers is a merely ordinal measure of the overall quality

of her performance in those exams, since the difference in quality between, for instance, the first- and second-ranked students might be far greater than, for instance, the difference in quality the second- and third-ranked students. Thus, exam rank is not an interval-scale measure of the quality of a student's exam performance, even though it is an ordinal measure of this.

Having reviewed this distinction between interval-scale and merely ordinal functions, I can now explain that there is in fact a close conceptual connection between this distinction and the NBS scale invariance axiom. Roughly speaking, Nash's scale invariance axiom tells us that it makes sense to apply the NBS only to bargaining problems wherein all of our bargainers' utility functions are interval-scale. Thus (on the flipside), it would *not* make sense to apply the NBS to any bargaining problems wherein one or more of our bargainers' utility functions are merely ordinal preference orderings over the possible outcomes.

The NBS is actually far from unique in this respect, since virtually all extant formal bargaining solutions are only applicable under the assumption that our utility representations are interval-scale.⁵ In §7.1.5 below, I will discuss what this means for the IMM approach to moral uncertainty.

Before turning to that discussion, however, I want to first of all take some time to explain exactly *why* Nash's scale invariance axiom tells us that we should apply the NBS only under the assumption of interval-scale utility representations. Explaining this conceptual connection will be my focus for

⁵Thomson 1994; Vanderschraaf 2023.

the entirety of the next subsection (§7.1.4), and so readers who are not interested in this discussion may wish to skip ahead now to §7.1.5.

7.1.4 Transformations

According to Nash's scale invariance axiom, 'any positive affine rescaling of any bargainers' utility functions should not alter the bargaining solution.' So, I shall begin this subsection by explaining what a 'positive affine' transformation is.

For any given pair of utility functions u_i and v_i , v_i is a positive affine rescaling of u_i iff $v_i := \alpha u_i + \beta$, where α and β are real-valued constants, and α is strictly positive. Thus, Nash's scale invariance axiom requires that replacing any bargainer's utility function u_i with any positive rescaling of the form $v_i := \alpha u_i + \beta$ (with $\alpha > 0$) should not alter the bargaining outcome selected by the NBS.

This property of the NBS is in fact quite easy to illustrate mathematically. Imagine, for the sake of simplicity, that there are only two bargainers, and that these two bargainers have the utility functions u_1 and u_2 respectively. Thus, under these conditions, any given outcome A will be an NBS iff setting $a = A$ maximises

$$[u_1(a) - d_1] \times [u_2(a) - d_2] \tag{7.1}$$

Now imagine, however, that we rescale the first bargainer's utility function,

like so:

$$v_1 := 10u_1 + 5$$

Thus any given outcome A will be an NBS under this rescaling iff setting $a = A$ maximises

$$[10u_1(a) + 5 - 10d_1 - 5] \times [u_2(a) - d_2] \quad (7.2)$$

But of course, this expression can simply be rewritten as

$$10 \times [u_1(a) - d_1] \times [u_2(a) - d_2] \quad (7.3)$$

which is simply ten times expression (7.1). Thus, any given choice of a maximises expression (7.2) iff that choice of a also maximises expression (7.1). Hence, rescaling u_1 to v_1 does not at all alter the NBS.

Now, notice that although Nash's scale invariance axiom does guarantee that the NBS will be invariant up to any given positive *affine* utility transformation, on the other hand it does *not* guarantee that the NBS will always be invariant under any positive *monotonic* utility transformation. For any given pair of utility functions u_i and v_i , v_i is a positive monotonic rescaling of u_i IFF: for any given pair of possible outcomes a and b , $v_i(a) > v_i(b)$ iff $u_i(a) > u_i(b)$.⁶ Hence, all possible positive affine transformations must also be positive monotonic, but some possible positive monotonic transfor-

⁶Or equivalently, v_i is a positive monotonic rescaling of u_i iff $v_i := f \circ u_i$, where f is some strictly increasing function.

mations are not also positive affine. For instance, suppose that $v_i := u_i^3$. Then it follows from this definition that v_i is a positive monotonic rescaling of u_i , because $u_i^3(a) > u_i^3(b)$ iff $u_i(a) > u_i(b)$. However, this is clearly not a positive *affine* rescaling of u_i , since v_i is not of the required form $\alpha u_i + \beta$.

Once again, it is quite easy to illustrate mathematically that the NBS need not always be invariant under positive monotonic utility transformations. As before, imagine for the sake of simplicity that there are only two bargainers, and that these two bargainers have the utility functions u_1 and u_2 respectively. Thus, under these conditions, any given outcome A will be an NBS iff setting $a = A$ maximises

$$[u_1(a) - d_1] \times [u_2(a) - d_2] \quad (7.1)$$

Now imagine, however, that we rescale the first bargainer's utility function by cubing it:

$$v_1 := u_1^3$$

Thus any given outcome A will be an NBS under this rescaling iff setting $a = A$ maximises

$$[u_1(a)^3 - d_1^3] \times [u_2(a) - d_2] \quad (7.4)$$

However, there is clearly no reason to think that expressions (7.1) and (7.4) should both be maximised by the same value of a . Hence, rescaling u_1 to v_1 could potentially alter which outcome is selected as the NBS.

Therefore, although the NBS is guaranteed to be invariant up to any positive affine utility transformations, by contrast it is not guaranteed to be invariant under all possible positive monotonic transformations. Thus, in light of this result, it makes sense to apply the NBS only to those bargaining problem in which each of the bargainer's utility functions represents something *more* than bargainer's preference *ordering* over the various possible outcomes. After all, if the ordering of all possible bargaining outcomes ranked from lowest to highest by their value according to any given utility function u_i is stipulated to be identical to some bargainer i 's preference ordering over all of these possible outcomes, then clearly the same must also hold true for any positive monotonic transformations v_i of u_i , since positive monotonic transformations are all 'order-preserving' by definition. Thus, if u_i does not represent anything over and above the bargainer i 's preference ordering over possible outcomes, there would be no reason to use u_i for this purpose rather than v_i . However, we already know that the choice between u_i and its positive monotonic transformation v_i *could* perhaps make a difference to the NBS for this particular bargaining problem. Thus, if we try to apply the NBS to any given bargaining problem wherein some of the bargainers have utility functions which only represent those bargainers' preference orderings over options, then under these conditions the NBS's output will be to at least some extent contingent upon certain totally arbitrary choices concerning which utility functions were used to represent those bargainers' preference orderings.

Clearly, then, it makes sense to apply the NBS only to bargaining problems in which each of the bargainers' utility functions represents something more than just the bargainers' preference orderings over the various possible outcomes. More specifically, Nash's scale invariance axiom tells us that we should apply the NBS only to bargaining problems wherein each of the bargainers' utility functions encodes some piece of information about that bargainer's preferences which could still be decoded from any given positive affine transformation of that utility function, but which could *not* be reliably decoded from any given merely positive monotonic transformation of that original utility function.

Now, as it happens this condition is in fact equivalent to the condition in §7.1.3 above stating that we can non-arbitrarily apply the NBS to any given bargaining problem only if all of our bargainers' utility representations are interval-scale. This follows from the fact that any given function u will interval-scale measure any given variable X iff any given positive affine rescaling $v := \alpha u + \beta$ of u would also be an interval-scale measure of the variable X . After all, if we assume that any given 1-unit increase in the value of u always represents some fixed increase of x in the variable X , then it follows that any given 1-unit increase in the value v must therefore always represent a fixed increase of $\frac{x}{\alpha}$ in the variable X . For example, °F is a positive affine transformation of °C, since ${}^{\circ}F := \frac{9}{5}{}^{\circ}C + 32$. Thus, any given 1-unit increase in °F always represent a fixed amount of extra heat which is $\frac{5}{9}$ the size of the fixed amount of extra heat that is represented by any given 1-unit increase

in °C.

Similarly, any given function u will at least ordinally measure any given variable X iff any given positive monotonic rescaling v of u would also at least ordinally measure the variable X . After all, if v is a positive monotonic rescaling of u , then by definition $u(a) > u(b)$ iff $v(a) > v(b)$, and so u and v will both represent exactly the same ordering over the domain of objects under evaluation. For example, exam rank at least ordinally measures the overall quality of my students' exam performances iff the same can also be said of, for example, *cubed* exam rank, since my exam rank will be higher than yours iff the same can also be said of our cubed exam ranks.

Thus, in light of these equivalences, Nash's scale invariance axiom tells us that we can non-arbitrarily apply the NBS only to bargaining problems wherein all of our bargainers' utility functions interval-scale represent something or other about those bargainers' preferences. Hence, I shall now turn my attention to the question of whether all possible IMM bargaining models can meet this condition.

7.1.5 Interval scales

I have just demonstrated that we can apply the NBS to bargaining problems between our IMM theory representatives only if we can assign to each of these theory representatives some utility function which interval-scale represents

something about her preferences.⁷ Unfortunately, however, there seems to be no compelling reason for us to assume that we could assign to each and every possible theory representative some utility function which would interval-scale represent some feature of their preferences. On the contrary, the set of all possible first-order moral theories is simply much too heterogeneous for this to count as a plausible assumption.

If this pessimistic view is correct, then we IMM theorists will face the difficult challenge of having to develop an alternative approach to formal bargaining which need *not* require interval-scale representations. I will offer a response to this challenge in §7.2 below. Before that, however, I want to first of all explain in greater detail why it would be implausible for us to suppose that each and every possible theory representative has interval-scale representable preferences.

First of all, in the economics literature on bargaining, the standard approach to ensuring the NBS interval-scale representation condition will be satisfiable is to adopt the assumption that every bargainer has von Neumann-Morgenstern (henceforth: ‘vNM’) preferences. That is, the economists tend to assume that every bargainer has some utility function whose *expected value* she will prefer to maximise in all possible circumstances involving any descriptive uncertainty. Granted this standard assumption, every bargainer’s utility

⁷More specifically, we need to be able to assume that there exists a certain determinable D such – holding this D *fixed* – we can assign each of our theory representatives some utility function which interval-scale represents the determinate of D which is instantiated by this representative’s preferences.

function will interval-scale represent that bargainer's *risk attitude* regarding descriptive uncertainty, with any given unit increase in the bargainer's vNM utility for any given outcome A always representing an increase of a certain fixed amount in the extent to which increasing the probability of that outcome A occurring would thereby improve the preferability of a descriptively risky lottery according to our particular bargainer.

Unfortunately, however, one simply cannot plausibly assume that all of the theory representatives for any given morally uncertain decision maker will all have vNM preferences. Admittedly, the orthodox view in decision theory says that expected utility maximisation is a requirement of practical rationality.⁸ But, at least some ethicists have argued that the best versions of, for instance, nonconsequentialist morality need not be vNM.⁹

Of course, many of us find ourselves unconvinced by any of the arguments against expected utility maximization. However, this fact notwithstanding, it would nonetheless strike me as highly implausible to assume that every reasonable moral decision maker must have absolutely *zero* credence in any first-order moral theory whose theory representative's preferences would be non-vNM. Hence, if we wish to develop an NBS version of IMM that could be

⁸Steele and Stefánsson 2020; Briggs 2023. (This is primarily because several important representation theorems demonstrate that any agent whose preferences satisfy certain sets of attractive axioms must therefore prefer to maximise expected utility; one can also defend expected utility maximisation by recourse to the Law of Large Numbers.)

⁹For example, Sergio Tenenbaum has argued that deontologists ought to reject expected value maximisation, since “deontological rules, prohibitions and permissions apply primarily to intentional acts; [and] risk changes the nature of an [intentional] act, not the probability that the same act will be performed” (Tenenbaum 2017, p. 675).

applicable to all reasonable moral uncertain decision maker, then we would have to have some other reason to think that all possible theory representatives must have interval-scale representable preferences.

One initially attractive possible response to this challenge might begin with the suggestion that our interval-scale utility function for any given theory representative ought to be a measure of the *strength* or *intensity* of the theory representatives' preferences for and against her decision maker's possible options. Sadly, however, I will now argue that it is highly doubtful whether we can legitimately assume that all possible theory representatives will have preferences over possible options whose 'strengths' or 'intensities' can be interval-scale represented by real-valued utility functions.

The strength of any given theory representative's preference for or against any possible option should plausibly just be determined by the *choiceworthiness* of that option according to the corresponding moral theory – where (as in §1.2 above) we say that the 'choiceworthiness' of any given option A according to any given moral theory T is just a measure of the strength of our decision maker's all-things-considered moral reason in favour of choosing A according to T. Now, under this assumption there will certainly exist *some* moral theory representatives whose preference strengths can be measured by interval-scale utility functions. Unfortunately, however, there will also exist many other representatives whose preference strengths *cannot* be modelled in this way. I will now present one example from each of these two categories.

Let me begin with one example of a first-order moral theory which implies

that the strengths of one's all-things-considered moral reasons can always be measured on an interval scale of choiceworthiness. According to certain consequentialist theories of morality, the strengths of one's moral reasons can always be interval-scale measured by a function from options to real numbers, which for any given option A is simply defined as the total moral (or 'impersonal') value of the state of the world which would eventuate if our decision maker selected option A. For the sake of concreteness, this 'moral value' function could simply be defined as the sum total of everybody's levels of well-being. In which case, we would clearly just be dealing with one possible version of utilitarianism, according to which a unit increase in the total wellbeing induced by any given option A will always represent an increase of a certain fixed amount in the strength of our decision maker's all-things-considered moral reason in favour of choosing the option A.

Moreover, it strikes me as highly plausible to suppose that this moral theory should have an IMM representative the strength of whose preference for or against any given option A is exactly equal to the strength of the moral reason for or against choosing A according to this particular version of utilitarianism. Thus, a unit increase in the total wellbeing induced by any given option A will always represent an increase of a certain fixed amount in the strength of this theory representative's preference for or against our decision maker choosing the option A. Hence, this is one example of a theory representative the strength of intensities of whose preferences can be measured on an interval-scale utility function.

Unfortunately, however, some other possible theory representatives will correspond to moral theories according to which the strengths of all-things-considered moral reasons cannot always be measured on some interval scale of choiceworthiness. For instance, according to one possible version of Kantian deontology, perhaps murder is less choiceworthy than lying, which is in turn less choiceworthy than failing to aid someone in need. But furthermore, perhaps this version of Kantian deontology also implies that it actually makes no sense to ask by what *amount* the strength of my moral reason against murdering differs from the strength of my moral reason against telling lies.¹⁰ In other words, this possible version of Kantianism will imply that the strengths of all-things-considered moral reasons can only ever be measured against a merely *ordinal* choiceworthiness scale, which would simply rank moral reasons from weakest to strongest – without attaching any ‘amounts’ to these strength differences. According to one possible such merely ordinal choiceworthiness scale, murder has choiceworthiness -2 , lying has choiceworthiness -1 , and failing to aid has choiceworthiness 0 . However, it would be just as legitimate to instead use a cubed version of this ordinal scale, on which murder would now have choiceworthiness -8 , lying would still have choiceworthiness -1 , and failing to aid would also still have choiceworthiness 0 .

Of course, both of these choiceworthiness scales should also ordinally measure the preferences over options of any adequate IMM theory representative for this Kantian moral theory. However, the information represented by these

¹⁰I borrow this example from MacAskill, Bykvist and Ord 2020, p. 60.

merely ordinal utility functions does not yet allow us to apply the NBS to any bargaining problems involving a representative for this version of Kantianism. (After all, I pointed out in §7.1.3 above that the solution outcomes selected by the NBS are unlikely to be invariant to positive monotonic but non-affine utility transformations like cubing.)

If we do in fact want to apply the NBS to any given bargaining problem involving a theory representative with ordinal preferences corresponding to this particular version of Kantianism, then we would first of all have to attach precise *amounts* to the differences in strength between the various ordinal preference levels into which this representative would sort all of the possible options for her decision maker. Or equivalently, we would have to stipulate that a certain class of representationally-equivalent interval-scale utility functions would all properly represent the strengths or intensities of the theory representative's preferences over the possible options. After that, we could then use any of these interval-scale utility functions to calculate some determinate NBS for our bargaining problem.

Unfortunately, however, there is no compelling reason to think that there could exist any principled way to choose exactly *which* set of amounts should be attached to the differences in strength between the different possible ordinal preference levels. In other words, it is impossible to imagine any principled way of choosing which class of representationally-equivalent interval-scale utility functions we should take to represent the strengths or intensities of our theory representative's preferences over all of the possible options.

After all, remember that the particular version of Kantianism corresponding to this representative implied that the strengths of our decision maker's all-things-considered moral reasons could not be represented by any given class of interval-scale choiceworthiness functions. Thus, I cannot imagine anything at all that we could possibly advert to as a principled reason for choosing any one particular utility interval-scale as opposed to all of its competitors.

Overall, then, we have not yet found any good reason to think that every possible moral theory representative will have interval-scale representable preferences. At the beginning of this subsection, I argued that the risk attitudes of all possible representatives cannot always be measured by interval-scale vNM utility functions.¹¹ Then, after that, I argued that the preference strengths or intensities of all possible theory representatives likewise cannot always be represented by interval-scale utility functions. Thus, we now face the challenge of developing an alternative approach to IMM bargaining which *will* be applicable to any given morally uncertain decision maker.

7.2 Subvaluation

Just a moment ago, I argued that certain moral theories need not supply us with any principled grounds for attaching any particular amounts to the

¹¹And actually, even if these risk attitudes could always be represented by interval-scale vNM utility functions, I would still be reluctant to incorporate those vNM utility functions into the IMM theory of appropriateness, because I worry that this would inappropriately disadvantage risk-averse theories in some choice situations that do not involve descriptive uncertainty (see Lloyd 2022, §2.3).

differences in strength between the various different ordinal preference levels into which these theories' representatives would sort all of our decision maker's possible options. Or, to put the same point slightly differently: I argued that IF we wish for every possible moral theory to be attached to one specific theory representative whose preferences over all of the possible options have strength differences that are interval-scale measurable, THEN unfortunately some of these representatives must thereby have preferences whose relative strengths are *underdetermined* by the moral theories to which they correspond. Hence, these kinds of moral theories will sometimes underdetermine exactly how their corresponding IMM representatives should want to bargain with each other.

Fortuitously, this reframing in terms of the notion of 'underdetermination' allows us to be reminded of an analogous problem for the IMM project to which I have already proposed a solution. More specifically, I argued in §6 above that we can describe certain theory representatives as having preferences which are *underdetermined* by their corresponding moral theories in some cases where these moral theories incorporate *prerogatives*. In that earlier discussion, I argued that we IMM theorists should handle this problem of underdetermination by adopting a certain 'subvaluationist' account of appropriateness. Suppose for the sake of simplicity that our decision maker faces a certain resource-division and 'constant returns to scale' choice situation Φ , and suppose that she is certain to never encounter any other choice situations in which any of the moral theories in which she has credence disagree

with each other. (As always, this assumption rules out any potential for intertemporal trades or contracts.) Moreover, I have stipulated that a preference assignment counts as *legitimate* for any given theory representative iff this representative having these assigned preferences would be compatible with her wishing for our decision maker to follow the directives of the moral theory corresponding to this representative. Then, under these conditions, my subvaluationist response to prerogatives would imply that any given option ϕ is a maximally appropriate response to Φ iff there exists *at least one* set of legitimate preference assignments under which our IMM theory representatives would jointly instruct our decision maker to choose ϕ in Φ .¹²

By analogy with this subvaluationist response to the problem of prerogatives, I now want to suggest that we should also adopt a subvaluationist response to the problem of merely ordinal moral theories. Once again, suppose for the sake of simplicity that our decision maker faces a certain resource-division and ‘constant returns to scale’ choice situation Φ , and suppose that she is certain to never encounter any other choice situations in which any of the moral theories in which she has credence disagree with each other. Now, however, let us stipulate that a preference assignment counts as *legitimate** for any given theory representative iff it assigns this theory representative preferences with interval-scale measurable strengths which are compatible with the directives of the moral theory corresponding to this representative.

¹²For a more general formulation of this version of IMM, see footnote 5 in §6 above.

Then, under these conditions, my new subvaluationist version of IMM would imply that any given option ϕ is a maximally appropriate response to Φ iff there exists *at least one* set of legitimate* preference assignments under which ϕ would be a Nash bargaining solution of our IMM bargaining model for Φ .¹³

If our morally uncertain decision maker has positive credence only in moral theories which imply that the strengths of her moral reasons can always be represented by interval-scale choiceworthiness functions – call these the *interval-scale* moral theories – then my new subvaluationist version of IMM will imply that ϕ is a maximally appropriate response to Φ iff ϕ is an NBS of our IMM bargaining model for Φ under *the* unique set of preference assignments which would specify that any given theory representative's preference for or against any given option A is exactly equal to the strength

¹³More generally (recall §5.6 above),

IF

- Ψ is the set of choice situations that our decision maker has confronted thus far in her lifetime;
- ψ is the set of choices that she has made in those situations;
- and Φ is the set of circumstances that our decision maker faces in deciding which plan to follow for the remainder of her lifetime

THEN any given course of action ϕ is a maximally appropriate response to Φ iff there exists at least one set of legitimate* IMM preference assignments under which ϕ minimises the shortfall from rendering every theory representative at worst indifferent (given their current descriptive credences) between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) some Nash bargaining solution of an IMM bargaining model of this sequence in which all of the decision rights are initially split between the theory representatives in proportion to our decision maker's (current) credences in the corresponding moral theories.

of the moral reason for or against choosing A according to the moral theory corresponding to this representative.

For example, recall the **Three Charities** choice situation from §2.4 above, in which some philanthropist has 50% credence in each of two moral theories T_1 and T_2 , and needs to decide how to distribute her fortune between three charities A, B, and C. Furthermore, suppose that both T_1 and T_2 are interval-scale moral theories. More particularly, suppose that according to T_1 , the strengths of the philanthropist's all-things-considered moral reasons can be interval-scale represented by the choiceworthiness function $CW_1(a, b, c)$, which assigns $10a + 9b + c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$ between the three charities A, B, and C. Moreover, suppose that according to T_2 , the strengths of our philanthropist's all-things-considered moral reasons can be represented by the choiceworthiness function $CW_2(a, b, c)$, which assigns $a + 9b + 10c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$. Finally, assume for the sake of simplicity that our philanthropist is certain to never encounter any other choice situations in which T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to this particular version of **Three Charities**.

According to my preferred subvaluationist version of IMM, any given donation distribution $\langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ is a maximally appropriate response to this particular version of **Three Charities** iff $\langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ is an NBS of the unique IMM bargaining model for this choice situation wherein the strengths of R_1 and R_2 's preferences over possible donation distributions are

interval-scale represented by CW_1 and CW_2 respectively.

Now, in order to calculate the NBS for this particular bargaining model, we will first of all need to calculate R_1 and R_2 's disagreement utilities d_1 and d_2 . Recall from §7.1.2 above that my basic idea for calculating disagreement utilities was that each theory representative's disagreement utility in any given sequence of choice situations if none of these representatives could agree to any trades or contracts. Hence, in the particular version of the **Three Charities** choice situation which I have just described, R_1 and R_2 's disagreement utilities should just be the utilities that R_1 and R_2 would have if R_1 were to use her initial 50% endowment of control rights to instruct our philanthropist to donate 50% of her inheritance to R_1 's favoured charity A, and if at the same time R_2 were to use her initial 50% endowment of control rights to instruct our philanthropist to donate 50% of her inheritance to R_2 's favoured charity C. In other words, d_1 should be calculated as

$$CW_1(50, 0, 50) = (10 \times 50) + (9 \times 0) + (50) = 550$$

and d_2 should be calculated as

$$CW_2(50, 0, 50) = (50) + (9 \times 0) + (10 \times 50) = 550$$

Thus, R_1 and R_2 should both have disagreement utilities of 550 in this particular version of **Three Charities**. Hence, my preferred version of IMM will imply that any given donation distribution $\langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ is a maximally ap-

propriate response to this particular version of **Three Charities** iff setting $\langle a\%, b\%, c\% \rangle = \langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ maximises the Nash maximand

$$(10a + 9b + c - 550) \times (a + 9b + 10c - 550) \quad (7.5)$$

subject to the constraints $10a + 9b + c \geq 550$ and $a + 9b + 10c \geq 550$.

As it turns out, expression 7.5 will be uniquely maximised when $\langle a, b, c \rangle = \langle 0, 100, 0 \rangle$. Thus, my preferred version of IMM would imply that our philanthropist donating her entire fortune to charity B is the only maximally appropriate possible response to this particular version of **Three Charities**. Fortunately, this implication strikes me as being highly plausible. Indeed, I have argued in §2.4 above that any plausible precisification of IMM must imply that it is most appropriate for our philanthropist to donate her entire fortune to charity B in this particular ‘identical bargaining positions’ version of **Three Charities**. Hence, my preferred version of IMM honours my intuitions regarding this particular choice situation.

At this point, perhaps it is worth mentioning that there will also exist at least some possible interval-scale variations on the **Three Charities** choice situation in which my preferred version of IMM would *not* imply that it is most appropriate for our philanthropist to donate her entire fortune to charity B. For the sake of illustration, let us consider one particular ‘non-identical bargaining positions’ variation on **Three Charities**. Imagine that in this variant choice situation, the strengths of R_1 ’s preferences can still be

interval-scale represented by $u_1(a, b, c) := 10a + 9b + c$. By contrast, however, let us now imagine that the strengths of R_2 's preferences could instead be interval-scale represented by $u_2(a, b, c) := a + \underline{6}b + 10c$ (with a '6' coefficient on the b instead of a '9').

Now, under these new conditions, R_1 and R_2 should both still have a disagreement utility of 550. Hence, my preferred subvaluationist version of IMM will imply that any given donation distribution $\langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ is a maximally appropriate response to this particular variation on **Three Charities** iff setting $\langle a\%, b\%, c\% \rangle = \langle \mathcal{A}\%, \mathcal{B}\%, \mathcal{C}\% \rangle$ maximises the Nash maximand

$$(10a + 9b + c - 550) \times (a + 6b + 10c - 550) \quad (7.6)$$

subject to the constraints $10a + 9b + c \geq 550$ and $a + 6b + 10c \geq 550$.

As it turns out, expression 7.6 will be uniquely maximised when $\langle a, b, c \rangle = \langle 0, 84.37, 15.63 \rangle$. Thus, my preferred version of IMM would imply that our philanthropist splitting her fortune 84.37:15.63 between the two charities B and C would be the only maximally appropriate response to this particular variation on **Three Charities**. This result would be explained by the fact that our compromise charity B is very nearly as good as the best available charity A according to the moral theory T_1 , whereas (on the other hand) this compromise charity B is not nearly as good as the best available charity C according to the moral theory T_2 . Hence, the theory representative R_1 would – relatively speaking – gain much more than the theory representative R_2

would from, for instance, a contract to replace the ‘default’ 50:50 distribution between charities A and C with our philanthropist donating her entire fortune to the compromise charity B. Thus, under these conditions, my preferred version of IMM would imply that it is most appropriate for our philanthropist to split her fortune 84.37:15.63 between B and C. Fortunately, this strikes me as a plausible response to this particular variation on the **Three Charities** choice situation.

In any case, this variation on **Three Charities** is also another example of a choice situation in which my preferred version of IMM implies that only one possible option is maximally appropriate. However, there are also many other possible choice situations in which my preferred version of IMM will imply that many possible options can all be maximally appropriate. After all, for any given uncertain decision maker who has positive credence in at least one moral theory which implies that the strengths of moral reasons cannot always be represented by interval-scale choiceworthiness functions – call these the *ordinal* moral theories – my new subvaluationist version of IMM will imply that ϕ is a maximally appropriate response to Φ iff ϕ is an NBS of our IMM bargaining model for Φ under at least one from a whole *range* of equally legitimate potential preference assignment sets. In this respect, my new subvaluationist approach in some sense injects a new degree of *latitude* into the appropriateness recommendations of IMM in cases wherein the strengths of some representatives’ preferences cannot be fully determined by their corresponding moral theories. In these kinds of cases, a

decision maker's credence distribution over moral theories need not always 'pin down' only one course of action as her maximally appropriate response to any given set of circumstances.

For example, consider the following variation on the **Three Charities** choice situation:

Ordinal Bargaining: some philanthropist is deciding where to donate her fortune. She faces a choice between three charities: A, B, and C. Suppose that this philanthropist has 50% credence in the moral theory T_1 , and 50% credence in T_2 . Suppose that both of these two moral theories are ordinal, and that according to T_1 , the merely ordinal strengths of the philanthropist's all-things-considered moral reasons can be represented by the choiceworthiness function $CW_1(a, b, c)$, which assigns $10a + 9b + c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$ between the three charities A, B, and C. Moreover, according to T_2 , the merely ordinal strengths of our philanthropist's all-things-considered moral reasons can be represented by the choiceworthiness function $CW_2(a, b, c)$, which assigns $a + 6b + 10c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$.¹⁴

And as we always do, let us assume for the sake of simplicity that our philanthropist is certain to never encounter any other choice situations in which

¹⁴Note that R_1 and R_2 's bargaining positions are *not* identical here, since the preferences ordinally represented by $a + 6b + 10c$ *cannot* also be ordinally represented by $a + 9b + 10c$.

T_1 and T_2 disagree, thus ruling out the potential for any gains from trade or contract involving other scenarios in addition to **Ordinal Bargaining**.

Just as in **Three Charities**, there are clearly gains from contract available in **Ordinal Bargaining**. For instance, R_1 and R_2 would both gain relative to the no-contract outcome if they together agreed to instruct our philanthropist to donate to charity B.¹⁵ However, in **Ordinal Bargaining** T_1 and T_2 are also now stipulated to be moral theories that reject interval-scale representations of choiceworthiness. Hence, my new subvaluationist version of IMM implies that a whole range of possible donation distributions are all maximally appropriate possible responses to this choice situation.

For instance, our philanthropist splitting her donations 84.37:15.63 between the two charities B and C is one maximally appropriate possible response to **Ordinal Bargaining**, since this is the unique NBS of the IMM model of this choice situation wherein the strengths of R_1 and R_2 's preferences over possible donation distributions are now taken to be interval-scale represented by CW_1 and CW_2 respectively.¹⁶ However, another maximally appropriate possible response to **Ordinal Bargaining** would be for our philanthropist to split her donations 89.12:10.88 between B and C, since this is the unique NBS of the IMM model wherein the strengths of R_1 's preferences are instead taken to be interval-scale represented by CW_1 *cubed*, whereas –

¹⁵After all, $CW_1(0, 100, 0) = 900 > 550 = CW_1(50, 0, 50)$, and $CW_2(0, 100, 0) = 600 > 550 = CW_2(50, 0, 50)$.

¹⁶In other words, setting $\langle a\%, b\%, c\% \rangle = \langle 0, 84.37, 15.63 \rangle$ uniquely maximises $(10a + 9b + c - 550) \times (a + 6b + 10c - 550)$.

by contrast – the strengths of R_2 's preferences are once again taken to be interval-scale represented by CW_2 *simpliciter*.¹⁷ Moreover, yet another maximally appropriate possible response to **Ordinal Bargaining** would be for our philanthropist to split her donations 81.73:18.26 between B and C, since this is the unique NBS of the IMM model wherein the strengths of R_1 's preferences are now taken to be interval-scale represented by CW_1 *simpliciter*, whereas the strengths of R_2 's preferences now taken to be interval-scale represented by CW_2 *cubed*.¹⁸

All three of these maximally appropriate possible donation distributions would require our philanthropist to donate none of her fortune to charity A, most of her fortune to charity B, and the rest of it to charity C. This result can be explained by the fact that according to the moral theory T_1 , shifting money from the best available charity A over to the compromise charity B would not do much to alter the ordinal preferability ranking of any given donation distribution; whereas (by contrast) according to the moral theory T_2 , shifting money from the best available charity C over to the compromise charity B would much more substantially alter the ordinal preferability ranking of any given donation distribution. Hence, the theory representative R_1 would – relatively speaking – come many more ranks closer to her ideal donation distribution than the theory representative R_2 would under, for instance, a

¹⁷In other words, setting $\langle a\%, b\%, c\% \rangle = \langle 0, 89.12, 10.88 \rangle$ uniquely maximises $\left[(10a + 9b + c)^3 - 550^3 \right] \times [a + 6b + 10c - 550]$.

¹⁸In other words, setting $\langle a\%, b\%, c\% \rangle = \langle 0, 81.73, 18.26 \rangle$ uniquely maximises $[10a + 9b + c - 550] \times \left[(a + 6b + 10c)^3 - 550^3 \right]$.

contract to replace the ‘default’ 50:50 distribution between A and C with an instruction for our philanthropist to donate all of her fortune to charity B. Thus, under these conditions by preferred version of IMM implies that it is most appropriate for our philanthropist to split her fortune between the two charities B and C. Once again, this strikes me as a plausible result.

In any case, however, my main reason for introducing this **Ordinal Bargaining** choice situation was for it to serve as an illustration of the fact that my preferred subvaluationist version of IMM implies that in certain choice situations many possible options can all be maximally appropriate. More precisely, my preferred subvaluationist version of IMM has this implication in certain choice situations wherein the morally uncertain decision maker has positive credence in at least one merely ordinal moral theory.

Incorporating this element of latitude into my IMM response to these sorts of choice situations does not strike me as implausible. Actually, it is perhaps unsurprising for the best version of IMM to have some latitude or indeterminacy in at least some cases which involve ordinal moral theories. The basic idea behind IMM is that any given course of action ϕ should count as a maximally appropriate possible response to any given set of circumstances Φ iff all of the theory representatives in our IMM bargaining model for Φ could end up jointly instructing our decision maker to choose the course of action ϕ . Moreover, one important determinant of any given agent’s bargaining priorities in this IMM model will surely be the relative differences in the strengths of her preferences over the various possible bargaining outcomes.

But, these relative strengths will unfortunately be indeterminate for the IMM representatives of any possible ordinal moral theories. Thus, it should be unsurprising that in certain cases involving these ordinal moral theories, IMM should have to incorporate an element of latitude.

In this subsection, I have discussed in some detail how IMM should handle the problem of ordinal moral theories. Nonetheless, from now on I will for the sake of simplicity typically restrict our attention only to choice situations involving interval-scale moral theories. In all of these choice situations, we can of course safely ignore all of the subvaluationist complexities introduced into my preferred version of IMM, since these complexities are relevant only in certain choice situations which involve ordinal moral theories.

7.3 Shortfalls

According to my preferred subvaluationist version of IMM,

IF

- Ψ is the set of choice situations that our decision maker has confronted thus far in her lifetime;
- ψ is the set of choices that she has made in those situations;
- and Φ is the set of circumstances that our decision maker faces in deciding which plan to follow for the remainder of her lifetime

THEN any given course of action ϕ is a maximally appropriate response to Φ iff there exists at least one set of legitimate* IMM preference assignments under which ϕ minimises the shortfall from rendering every theory representative at worst indifferent (given their current descriptive credences) between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) some Nash bargaining solution of an IMM bargaining model of this sequence in which all of the decision rights are initially split between the theory representatives in proportion to our decision maker's (current) credences in the corresponding moral theories.

For certain specifications of Ψ , ψ , and Φ , there will exist some legitimate* IMM preference assignments under which some option ϕ would in fact render every theory representative *exactly* indifferent between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) some NBS of the IMM model for this sequence. For example, I demonstrated in §7.2 above that $\langle 0, 100, 0 \rangle$ is the unique such response when Φ is at least one possible version of the **Three Charities** choice situation.

However, for some other specifications of Ψ , ψ , and Φ , there might *not* exist any legitimate* IMM preference assignments under which some option ϕ would render every theory representative exactly indifferent between (a) repeating ψ -then- ϕ many times over in response to a repeated sequence of instances of Ψ -then- Φ , and (b) some NBS of the IMM model for this se-

quence. Indeed, I have already discussed two impediments to satisfying this indifference condition in two earlier chapters of this dissertation. Firstly, I demonstrated in §4.2.2 above that in certain circumstances involving past noncompliance, it can sometimes be impossible to now satisfy IMM's indifference desideratum. Furthermore, I also demonstrated in §5.4.4 above that it can likewise be impossible to satisfy this desideratum in certain circumstances involving discrete choice.

In order to handle circumstances of this sort, we will need to specify how we should measure the extent to which any given course of action ϕ would 'fall short of' rendering every theory representative at worst indifferent between (a) repeating ψ -then- ϕ – or ' $\chi_{\text{rep}\psi\phi}$ ' for short, and (b) any given NBS of our IMM model χ_{nbs} . Following my earlier suggestion from §4.2.2 above, we can split this problem up into two connected subproblems, *viz.*: (1) specifying how to measure the extent to which the course of action ϕ would fall short of rendering any particular theory representative at worst indifferent between $\chi_{\text{rep}\psi\phi}$ and χ_{nbs} , and (2) specifying how to aggregate these measurements for each individual theory into an overall measure of the extent to which ϕ would fall short of rendering *every* theory representative at worst indifferent between $\chi_{\text{rep}\psi\phi}$ and χ_{nbs} .

§4.2.3 above included a first-pass characterization of my preferred approach to measuring shortfalls from at worst indifference for any given theory representative. The main idea behind this first-pass statements was that we should measure the extent to which any given theory representative falls

short of having at least its due influence in terms of the extent to which this theory representative is closer to having its due influence than it is to having exactly *zero* influence. Thus, the extent to which $\chi_{\text{rep}\psi\phi}$ falls short of giving any particular theory representative R_i the (due) level of influence that this theory representative would have enjoyed under χ_{nbs} should be measured relative to the choiceworthiness difference according to R_i between (1) χ_{nbs} , and (2) the course of behaviour χ_{-i} that our decision maker would have been instructed to realise in our repeated sequence of instances of Ψ -then- Φ in a version of our IMM model for this sequence of choice situations in which R_i is not endowed with any control rights, and hence has *zero* influence over our decision maker's behaviour.

In other words, I suggest that the extent $\Lambda_i(\chi_{\text{rep}\psi\phi}, \chi_{\text{nbs}})$ to which ϕ falls short of rendering R_i at worst indifferent between $\chi_{\text{rep}\psi\phi}$ and any given χ_{nbs} should be measured by something like the relative choiceworthiness difference

$$\Lambda_i(\chi_{\text{rep}\psi\phi}, \chi_{\text{nbs}}) := \frac{CW_i(\chi_{\text{nbs}}) - CW_i(\chi_{\text{rep}\psi\phi})}{CW_i(\chi_{\text{nbs}}) - CW_i(\chi_{-i})}$$

where:

IF c denotes our decision maker's credence function over moral theories,

THEN χ_{-i} can be defined as the course of action that our IMM bargainers would instruct our decision maker to choose in response to the sequence of repeated instances of Ψ -then- Φ if her

credence function over moral theories were instead c_{-i} , where

$$c_{-i}(t) := \begin{cases} 0 & \text{if } t = T_i \\ \frac{c(t)}{1-c(T_i)} & \text{otherwise} \end{cases}$$

I have described this characterization of Λ_i as a *first-pass* statement of the main idea behind my preferred approach to measuring shortfalls from at worst indifference. And that is just as well, because this first-pass formula for Λ_i turns out to suffer from one important problem, *viz.* that its value need not be invariant under positive monotonic transformations of T_i 's choiceworthiness function CW_i .¹⁹ For example,

$$\frac{CW_i(\chi_{\text{nbs}})^3 - CW_i(\chi_{\text{rep}\psi\phi})^3}{CW_i(\chi_{\text{nbs}})^3 - CW_i(\chi_{-i})^3}$$

¹⁹Another potential problem with this first-pass definition of Λ_i lies in my definition of χ_{-i} as ‘the course of action that our IMM bargainers would instruct our decision maker to choose in response to the sequence of repeated instances of Ψ -then- Φ if her credence function over moral theories were instead c_{-i} .’ The problem with this definition of χ_{-i} is that there is nothing to guarantee that the alternative IMM bargaining model with credence c_{-i} will have only *one* unique NBS course of action. In some unusual set of circumstances, we can imagine IMM bargaining models with credences c_{-i} which have multiple NBSs. Thus, the definite description which I used to define χ_{-i} could sometimes suffer from a failure of presupposition.

For the sake of brevity, I will not discuss this problem any further in this dissertation – except to mention that its range of potential solutions will be closely analogous to the range of potential solutions to the problem of underdetermined disagreement utilities which I introduced in §7.1.2 above.

will often have a different value from that of

$$\Lambda_i(\chi_{\text{rep}\psi\phi}, \chi_{\text{nbs}}) = \frac{CW_i(\chi_{\text{nbs}}) - CW_i(\chi_{\text{rep}\psi\phi})}{CW_i(\chi_{\text{nbs}}) - CW_i(\chi_{-i})}$$

– despite the fact that cubing is a positive monotonic transformation. Thus, it would make sense to use this definition of the shortfall function Λ_i only in cases where the choiceworthiness function CW_i can be taken to represent something more than just T_i 's preference *ordering* over the various possible outcomes.

Fortunately, this problem with my proposed formula for Λ_i is quite easy to solve when placed in the context of my preferred subvaluationist version of IMM. Recall that according to this version of IMM, the appropriateness of any given course of action ϕ depends upon the extent to which ϕ falls short of rendering each theory representative R_i at worst indifferent between $\chi_{\text{rep}\psi\phi}$ and any given χ_{nbs} , in an IMM bargaining model where R_i 's preference strengths are always measurable on an interval-scale. Thus, we can simply stipulate that if u_i is an interval-scale representation of R_i 's preference strengths, then

$$\Lambda_i(\chi_{\text{rep}\psi\phi}, \chi_{\text{nbs}}) := \frac{u_i(\chi_{\text{nbs}}) - u_i(\chi_{\text{rep}\psi\phi})}{u_i(\chi_{\text{nbs}}) - u_i(\chi_{-i})}$$

This new definition guarantees that the values of the shortfall function $\Lambda_i(\chi_{\text{rep}\psi\phi}, \chi_{\text{nbs}})$ will be the same regardless of exactly which utility function u_i is used to interval-scale represent R_i 's preference strengths. This is because

my new formula for Λ_i is guaranteed to be invariant under all possible positive affine transformations of u_i . This result is quite easy to prove mathematically: if $\alpha > 0$, then clearly

$$\begin{aligned} \frac{\alpha u_i(\mathbf{x}_{\text{nbs}}) + \beta - \alpha u_i(\mathbf{x}_{\text{rep}\psi\phi}) - \beta}{\alpha u_i(\mathbf{x}_{\text{nbs}}) + \beta - \alpha u_i(\mathbf{x}_{-i}) - \beta} &= \frac{\alpha [u_i(\mathbf{x}_{\text{nbs}}) - u_i(\mathbf{x}_{\text{rep}\psi\phi})]}{\alpha [u_i(\mathbf{x}_{\text{nbs}}) - u_i(\mathbf{x}_{-i})]} \\ &= \frac{u_i(\mathbf{x}_{\text{nbs}}) - u_i(\mathbf{x}_{\text{rep}\psi\phi})}{u_i(\mathbf{x}_{\text{nbs}}) - u_i(\mathbf{x}_{-i})} \end{aligned}$$

Thus, as desired, the value of the shortfall function Λ_i will not depend upon our choice of which utility function u_i is used to interval-scale represent R_i 's preference strengths.

Chapter 8

Conclusion

8.1 Evaluation

In this dissertation, I have proposed and developed the IMM approach to moral uncertainty. Along the way, I have argued that this new approach has several important advantages over its competitors. In summary:

1. In §§2.2-2.3 above, I argued that in certain choice situations like **Philanthropy**, IMM's recommendations to compromise are more intuitively appealing than MEC, MFT and MFO's 'winner-take-all' recommendations.
2. In §2.4 above, I argued that IMM underwrites 'hedging' in choice situations like **Three Charities** – this is an advantage over MFT and MFO which IMM has in common with MEC.

3. In §2.5 above, I argued that IMM underwrites ‘stakes sensitivity’ in choice situations like **Double Distribution** – this is another advantage over MFT and MFO which IMM has in common with MEC.
4. In §6 above, I argued that IMM has the resources required to develop an attractive response to the problem of first-order moral prerogatives. It is unclear whether this is an advantage of IMM over MEC, because it is a topic of active philosophical debate whether or not MEC has the resources required to develop an attractive response to the problem of first-order moral prerogatives.¹
5. In chapter 7, I demonstrated that my preferred version of IMM can satisfy Nash’s scale invariance axiom. Thus, I have demonstrated that IMM need not require any intertheoretic choiceworthiness unit comparisons. This is another advantage of IMM over MEC (recall §1.2 above).
6. Finally, the fact that my preferred version of IMM respects scale invariance also implies that this version of IMM is *nonfanatical*. This is yet another advantage of IMM over MEC (recall §1.2 above).

These advantages of IMM are summarised in figure 8.1.

One potential objection to the IMM approach which I have developed over the course of this dissertation is that it is not particularly simple, parsimonious,

¹See Hedden 2016, §5.2.2; MacAskill, Bykvist and Ord 2020, pp. 51-3; Sung 2023; Kaczmarek and Lloyd forthcoming.

<i>desiderata:</i>	<i>IMM:</i>	<i>MEC:</i>	<i>MFT:</i>	<i>MFO:</i>
compromise instead of winner-takes-all, in e.g. Philanthropy	✓	✗	✗	✗
underwrites hedging, in e.g. Three Charities	✓	✓	✗	✗
underwrites stakes sensitivity, in e.g. Double Distribution	✓	✓	✗	✗
underwrites the value of moral information	✓	✓	✗	✗
plausible response to prerogatives	✓	?	✓	?
intertheoretic unit comparisons not required	✓	✗	✓	✓
non-fanatical	✓	✗	<i>n/a</i>	✓
best version is uncomplicated	✗	✗	✗	?
			<i>another disadvantage:</i> problem of theory individuation	<i>other disadvantages:</i> (1) cyclicity (2) dependence on irrelevant alternatives

Figure 8.1: Advantages and disadvantages of IMM, MEC, MFT and MFO

monious, or easy to apply in practice. However, I have two lines of response to this objection. Firstly, *tu quoque*, the best alternatives to IMM are also rather complicated. For instance, the modified version of MEC defended by MacAskill, Bykvist and Ord requires considerable ancillary apparatus to handle cases where choiceworthiness is not intertheoretically unit-comparable across every moral theory in which the decision maker has positive credence.² Indeed, MacAskill, Bykvist and Ord themselves report that the entire ancillary apparatus developed over several chapters of their book is in fact intended to handle only a proper subset of all of the logically possible forms of intertheoretic incomparability.³ Thus, an exhaustive extension of MacAskill, Bykvist and Ord's preferred version of MEC will presumably be even more complicated.

Secondly, many of us are open to the possibility that the correct first-order moral theory is rather complicated. In that case, why suppose that the correct theory of appropriateness is any less complex? At the first-order level, a distinction is often drawn between *standards of rightness* and *decision procedures*. A standard of rightness determines which actions are right and wrong. However, if a theory's standard of rightness is difficult to apply, then it will often be supplemented by a decision procedure: a rough heuristic for determining rightness and wrongness that a decision maker can feasibly be guided by in the real world. Similarly, I suggest that advocates of IMM

²MacAskill, Bykvist and Ord 2020.

³MacAskill, Bykvist and Ord 2020, pp. 5-9; cf. Tarsney 2021.

can attempt to develop a metanormative decision procedure that roughly approximates IMM by cutting a few corners. I hope to discuss this prospects for this sort of decision procedure in my future research.

8.2 Future research

In this dissertation, I have presented some of the most important elements that would be necessary for a complete development and defence of the IMM approach to moral uncertainty. However, there are also many other important topics which I have lacked the time or space to cover in this dissertation. I will now provide a (long) list of some of these topics which I hope to discuss in my future work on IMM.

- In §3 above, I suggested that if our morally uncertain decision maker also happens to be descriptively uncertain, then our IMM theory representatives should directly ‘inherit’ this decision maker’s descriptive uncertainty. This strikes me as an entirely plausible assumption in cases wherein our decision maker’s moral uncertainty is probabilistically independent of her descriptive uncertainty. However, in any cases wherein our decision maker’s moral and descriptive credences are probabilistically *dependent* upon each other, it will be more plausible to suppose that each theory representative’s descriptive credence distribution should be derived from the decision maker’s by conditionalising upon the truth of the moral theory corresponding to this particular

representative. I hope to develop and defend this idea in future work.

- Descriptive uncertainties also introduce another important complication, *viz.* the fact that certain first-order moral theories are ‘*objectivist*,’ in the sense that these moral theories imply that any given action’s choiceworthiness depends upon that action’s *actual* features, regardless of whether or not our decision maker knows or believes that this action will have those features. It is not obvious how a theory of appropriateness like IMM should handle these objectivist moral theories.
- In chapter 4 of this dissertation, I discussed several topics pertaining to intertemporal bargaining. However, one topic not discussed in this dissertation is the question of how IMM should handle choice situations in which our decision maker has the opportunity to gain new ‘moral information’ which would improve the accuracy of her moral credence distribution. In future work, I will argue that IMM can provide an account of the value of moral information. This is yet another attractive feature of IMM that has previously been regarded as a distinctive advantage of MEC in particular.
- Another topic that I hope to discuss in future research relates to my discussion of discrete choice in chapter 5 of this dissertation. In that chapter, I argued that we IMM theorists should adopt the ‘social planning’ response to discrete choice situations. However, another possible

alternative to the ‘simple lottery’ approach to handling discrete choice has been suggested elsewhere in the philosophical literature on moral uncertainty and bargaining.⁴ In future work, I hope to demonstrate that this alternative response is inferior to my social planning proposal.

- All of the resource division choice situations that I have discussed in this dissertation have had ‘constant returns to scale.’ But in my future research, I hope to explain how IMM can handle choice situations in which the returns to scale need not always be constant. My basic idea here will be to argue that non-constant returns to scale can be well-handled by the ‘social planning’ version of IMM which I stated in full in §5.6 of this dissertation.
- Relatedly, none of the choice situations that I have discussed in this dissertation have been ‘coordination problems’ – in which one or more of our theory representatives preferences over how it might use its initial endowment of control rights are sensitive to how some or all of the *other* representatives decide to use their own endowments. Once again, I hope to argue in future research that these kinds of choice situations can be well-handled by the ‘social planning’ version of IMM.
- Furthermore, all of the choice situations that I have considered in this dissertation have been choice situations in which our decision maker has a *discrete* credence distribution over a finite set of possible moral

⁴Newberry and Ord 2021; Kaczmarek, Lloyd and Plant 2025; Lloyd 2025.

theories. However, many real-world decision makers are in fact likely to have *continuous* credence distributions. In future research, I hope to develop an IMM approach to handling these continuous credence distributions.

- In §6 above, I discussed how IMM should handle first-order moral theories which incorporate prerogatives. This discussion can be understood as IMM's first step along the road to engaging with the much broader problem of '*structural diversity*' amongst first-order moral theories.⁵ In future work, I hope to discuss how the subvaluationist response to prerogatives which I developed in §6 can be extended to handle other possible examples of structural diversity.
- Throughout this dissertation, I have been presenting IMM as a theory about which options are *maximally* appropriate in any given choice situation. However, it is natural to think that within the set of options which fall short of maximal appropriateness in any given choice situation, some options will be more or less inappropriate than others. In future research, I hope to demonstrate how we can vindicate this suggestion using the concept of 'shortfalls from due influence' which I developed in §§4.2.3, 4.2.4, and 7.3 of this dissertation.
- In chapter 7 above, I discussed how to incorporate the Nash bargaining solution into my preferred version of IMM. And in §7.2 above, I applied

⁵See Tarsney 2021.

the NBS to the **Three Charities** choice situation. Yet, in future research, I hope to discuss how the NBS can be applied to many other choice situations in addition to **Three Charities**.

- Relatedly, I also hope to be able to characterize much more systematically what the implications would be of my preferred subvaluationist response to merely ordinal moral theories. For example, I hope to be able to give a complete analytic characteristic of the range of donation distributions which would be appropriate in **Ordinal Bargaining**.
- In future research, I hope to develop a more sustained plan of attack for undermining the appeal of any rivals to IMM – most especially MEC. In particular, I hope to:
 - challenge some potential defences of intertheoretic unit comparisons;
 - discuss the problem of fanaticism;⁶
 - debunk MEC’s motivating idea that there is a close analogy between moral and descriptive uncertainty.
- On the other hand, one potential objection to IMM which I have not addressed in this dissertation would concern the fact that according to IMM, the appropriatenesses of our decision maker’s options at any given point in time depends upon how this decision maker has behaved

⁶Cf. Baker 2024

over the rest of her lifetime thus far, and will also depend upon the sequence of choice situations that this decision maker expects to confront for the remainder of her lifetime. In future work, I hope to argue that this is a bullet worth biting for IMM.

- Another large and important topic not discussed in this dissertation is the *metaethics* of ‘appropriateness.’ How exactly should we explicate what it means to say that an option is ‘appropriate’ or ‘inappropriate’? I hope to survey some of the potential responses to this question in my future work.
- Relatedly, I also hope to address ‘metanormative regress’ objections to the entire project of building appropriateness theories such as IMM.
- All of the example choice situations which I have discussed in this dissertation have taken the form of highly stylized hypothetical cases. But, in my future work I hope to discuss how IMM can be applied to real world choice situations covering topics including: career choice; vegetarianism; abortion; philanthropic organizations; and artificial intelligence alignment.
- In future work, I hope to expand upon the argument made in §8.1 above which said that objections to the complexity of IMM can be sidestepped by developing a heuristic decision procedure to approximate IMM.
- Finally, I hope to be able to say more in response to the question of

why we should even be interested in theories of appropriateness (such as IMM).

8.3 Coda

In this dissertation, I have developed and defended a new approach to the problem of moral uncertainty – a problem which has only recently begun to receive the level of philosophical attention it deserves. The IMM approach that I have developed here proceeded from an attractive motivating idea, and gave plausible recommendations in a wide range of cases. Of course, there is much work still to be done to develop all of the details of the IMM theory. Nonetheless, I hope to have at least convinced you here that the IMM research project offers a highly promising new response to the problem of moral uncertainty.

List of choice situations

- **Philanthropy (§1.3):** some philanthropist is deciding where to donate her fortune. She faces a choice between two charities: one that provides deworming pills to distant children, and a second that supports local soup kitchens. Suppose that this philanthropist has 60% credence in the moral theory T_1 , according to which she should donate as much as possible to deworming, and 40% credence in the moral theory T_2 , according to which she should donate as much as possible to soup kitchens.
- **Ninety-Nine Charities (§2.3):** some philanthropist is deciding where to donate her fortune. She faces a choice between ninety-nine different charities: C_1 through C_{99} . Suppose that this philanthropist has 1% credence in each of the ninety-eight different moral theories T_1 through T_{98} , and 2% credence in a final moral theory T_{99} . According to each moral theory T_n , the same-numbered charity C_n is the only charity that it would ever be choiceworthy for the philanthropist to donate to. In fact, each moral theory T_n implies that the choiceworthiness of

any possible donation distribution is linearly proportional exactly and only to the amount that the philanthropist donates to the corresponding charity C_n . (Hence, marginal choiceworthiness is non-diminishing according to every moral theory T_1 through T_{99} .)

- **Three Charities (§2.4):** some philanthropist is deciding where to donate her fortune. She faces a choice between three charities: A, B, and C. Suppose that this philanthropist has 50% credence in the moral theory T_1 , and 50% credence in T_2 . According to T_1 , it is highly choiceworthy to donate to A, almost as choiceworthy to donate to B, but scarcely choiceworthy at all to donate to C. Conversely, according to T_2 , it is highly choiceworthy to donate to C, almost as choiceworthy to donate to B, but scarcely choiceworthy at all to donate to A (illustrated in figure 2.3).
- **Double Distribution (§2.5):** some decision maker is choosing (i) where to donate her fortune, and (ii) what to do with her free time. This decision maker must distribute her fortune between two charities: the first of which provides deworming pills to distant children, and the second of which supports local soup kitchens. Similarly, our decision maker must also distribute her free time between two possible uses: the first of which is campaigning for nuclear disarmament, and the second of which is volunteering at a local orphanage. Suppose that this decision maker has, say, 50% credence in the moral theory T_1 , according to which

she should donate as much of her money as possible to deworming, and as much of her time as possible to nuclear disarmament. She also has 50% credence in the moral theory T_2 , according to which she should donate as much of her money as possible to soup kitchens, and as much of her time as possible to the local orphanage.

- **Risky Philanthropy (§3):** some philanthropist is deciding where to donate her fortune. She faces a choice between two options: (1) a charity that provides deworming pills to distant children; and (2) investing in a new social enterprise corporation. Suppose the philanthropist thinks there is some chance that the social enterprise corporation will benefit a large number of people in her local community. However, she also thinks there is some chance that it will fail to benefit anyone. This philanthropist has 60% credence in an impartialist moral theory T_1 according to which aiding distant strangers is *ceteris paribus* no less choiceworthy than aiding her local community; but she also has 40% credence in a partialist moral theory T_2 according to which aiding her local community is *ceteris paribus* somewhat more choiceworthy than aiding distant strangers.
- **Inheritance (§4.1):** some decision maker is choosing how to divide her free time between two possible uses: one of which is campaigning for nuclear disarmament, and the second of which is volunteering at a local orphanage. Furthermore, this decision maker is also certain

that within the next few months, she will inherit some money from her dying grandmother, which she will then have to distribute between two charities: one of which provides deworming pills to distant children, and the second of which supports local soup kitchens. As in **Double Distribution**, suppose that this decision maker has, say, 50% credence in the moral theory T_1 according to which she should spend as much of her free time as possible campaigning for nuclear disarmament, and should eventually donate as much of her inheritance as possible to the deworming charity. She also has 50% credence in the moral theory T_2 according to which she should spend as much of her free time as possible at the local orphanage, and should eventually donate as much of her inheritance as possible to the local soup kitchens.

- **Risky Inheritance (§4.1):** some decision maker is choosing how to divide her free time between disarmament campaigning and orphanage volunteering. Furthermore, this decision maker is also certain that within the next few months, she will inherit some money from her dying grandmother, which she will then have to distribute between deworming and soup kitchens. However, this decision maker is uncertain about how much money her grandmother will bequeath to her. Suppose that this decision maker has 99% credence in T_1 , which favours disarmament campaigning and deworming; with 1% credence in T_2 , which favours orphanage volunteering and soup kitchens.

- **Trolley (§5):** a runaway trolley is headed towards five people who are tied to the tracks. The decision maker in this scenario has three options:

- (i) pushing a heavyset man into the path of the trolley, intentionally killing him but saving the five;
- (ii) pulling a lever to redirect the trolley onto a side track where only two people are trapped, foreseeably killing them but saving the five;
- (iii) doing nothing, allowing the five to die.

Furthermore, suppose that this decision maker has 50% credence in a consequentialist moral theory T_1 , and 50% credence in a deontological moral theory T_2 . According to T_1 , whenever our decision maker confronts a choice situation like **Trolley**, it will be most choiceworthy to push the heavyset man, almost as choiceworthy to pull the lever, and highly unchoiceworthy to do nothing. Conversely, according to T_2 , whenever our decision maker confronts a choice situation like **Trolley**, it will be most choiceworthy to do nothing, almost as choiceworthy to pull the lever, and highly unchoiceworthy to push the heavyset man (illustrated in figure 5.1).

- **Special Obligation (§5.1):** some decision maker faces a choice between two possible options:
 - (i) averting a lesser harm from befalling the decision maker's parents;
 - (ii) averting a greater harm from befalling a stranger.

This decision maker has 90% credence in an impartialist moral theory T_1 according to which she is required to aid the stranger, but 10% credence in a partialist moral theory T_2 according to which she is required to aid her parents.

- **X (§5.4.1):** generic resource-division choice situation.
- **Y (§5.4.6):** generic discrete-choice situation.
- **Operations (§6):** some decision maker is deciding how to divide her life savings. This decision maker must distribute her savings between two possible uses: the first of which is helping to fund an operation for a sick close friend of hers, and the second of which is helping fund similar operations for several distant strangers. Suppose that this decision maker has 99% credence in some moral theory T_1 , and 1% in another moral theory T_2 . According to the theory T_1 , moral agents have a strong agent-centred prerogative over how they use their own savings. Hence, T_1 implies that any possible distribution of money between aiding the friend and aiding the strangers would be morally permissible under these conditions. By contrast, according to T_2 , our decision maker is morally required to use all of her life savings to fund operations for the distant strangers under these circumstances.
- **Dominance (§6):** some decision maker is choosing between only two possible options, A and B. Suppose that this decision maker has positive

credence in only two different moral theories – one of which is total-utilitarian, and the other is deontological. As it happens, options A and B would both produce exactly the same amount of total wellbeing, and so the utilitarian theory implies that these two options are equally choiceworthy. By contrast, however, the deontological theory implies that A is the only permissible option in this choice situation.

- **Free Time (§6):** our decision maker has one hour of free time this afternoon, and now faces a choice between only two uses for it: (1) visiting her parents for an hour; or (2) spending an hour volunteering at some local charity. Suppose that according to T_1 , our decision maker should spend her free time visiting her parents; whereas according to T_2 , she should spend her time volunteering at the local charity.
- **Ordinal Bargaining (§7.2):** some philanthropist is deciding where to donate her fortune. She faces a choice between three charities: A, B, and C. Suppose that this philanthropist has 50% credence in the moral theory T_1 , and 50% credence in T_2 . Suppose that both of these two moral theories are ordinal, and that according to T_1 , the merely ordinal strengths of the philanthropist's all-things-considered moral reasons can be represented by the choiceworthiness function $CW_1(a, b, c)$, which assigns $10a + 9b + c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$ between the three charities A, B, and C. Moreover, according to T_2 , the merely ordinal strengths of our philanthropist's

all-things-considered moral reasons can be represented by the choice-worthiness function $CW_2(a, b, c)$, which assigns $a + 6b + 10c$ to any given donation distribution $\langle a\%, b\%, c\% \rangle$.

References

Baker, Calvin. 2024. Expected choiceworthiness and fanaticism. *Philosophical Studies*, 181.5, 1237-56.

Briggs, R. A. 2023. Normative theories of rational choice: expected utility. In Edward N. Zalta and U. Nodelman (ed), *The Stanford Encyclopedia of Philosophy* (Fall 2023).

<URL: <https://plato.stanford.edu/archives/win2023/entries/rationality-normative-utility/>>.

Broome, John. 2012. *Climate Matters: Ethics in a Warming World* (New York, NY: W. W. Norton).

Bykvist, Krister. 2014. Evaluative uncertainty, environmental ethics, and consequentialism. Pp. 122-35 in Avram Hiller, Ramona Ilea and Leonard Kahn (eds), *Consequentialism and Environmental Ethics* (New York, NY: Routledge).

Bykvist, Krister. 2017. Moral uncertainty. *Philosophy Compass*, 12.3, e12408.

Cobreros, Pablo. 2013. Vagueness: subvaluationism. *Philosophy Com-*

pass, 8.5, 472-85.

Cohen, Haim, Nissan-Rozen, Ittay and Maril, Anat. 2024. Empirical evidence for moral Bayesianism. *Philosophical Psychology*, 37.4, 801-30.

Geyer, Jay. 2018. Moral uncertainty and moral culpability. *Utilitas*, 30.4, 399-416.

Goldman, Holly S. 1978. Doing the best one can. Pp. 185-214 in Alvin I. Goldman and Jaegwon Kim (eds), *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt* (Dordrecht: Springer).

Gracely, Edward J. 1996. On the comparability of judgments made by different ethical theories. *Metaphilosophy*, 27.3, 327-32.

Greaves, Hilary and Cotton-Barratt, Owen. 2024. A bargaining-theoretic approach to moral uncertainty. *Journal of Moral Philosophy*, 21.1-2, 127-69.

Greaves, Hilary and Ord, Toby. 2017. Moral uncertainty about population axiology. *Journal of Ethics and Social Philosophy*, 12.2, 135-67.

Gustafsson, Johan E. 2022. Second thoughts about my favourite theory. *Pacific Philosophical Quarterly*, 103.3, 448-70.

Gustafsson, Johan E. and Torpman, Olle. 2014. In defence of My Favourite Theory. *Pacific Philosophical Quarterly*, 95.2, 159-74.

Hedden, Brian. 2016. Does MITE make right? On decision-making under normative uncertainty. Pp. 120-28 in Russ Shafer-Landau (ed), *Oxford Studies in Metaethics*, volume 11 (Oxford: Oxford University Press).

Hudson, James L. 1989. Subjectivization in ethics. *American Philosophical*

ical Quarterly, 26.3, 221-9.

Kaczmarek, Patrick and Lloyd, Harry R. Forthcoming. Moral uncertainty, pure justifiers, and agent-centred options. *Australasian Journal of Philosophy*.

Kaczmarek, Patrick, Lloyd, Harry R., and Plant, Michael. 2025. Moral uncertainty, proportionality, and bargaining. *Ergo*, 12.44, 1142-71.

Karnofsky, Holden. December 13, 2016. Worldview diversification. Open Philanthropy, blog post.

<URL: <https://www.openphilanthropy.org/research/worldview-diversification/>>.

Karnofsky, Holden. January 26, 2018. Update on cause prioritization at Open Philanthropy. Open Philanthropy, blog post.

<URL: <https://www.openphilanthropy.org/research/update-on-cause-prioritization-at-open-philanthropy/>>.

Lloyd, Harry R. 2022. The property rights approach to moral uncertainty. Happier Lives Institute, working paper.

Lloyd, Harry R. 2025. Disagreement, AI alignment, and bargaining. *Philosophical Studies*, 182.7, 1757-87

Lloyd, Harry R. Forthcoming. Moral uncertainty and expected truthlikeness. *Synthese*.

Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences* (Oxford: Oxford University Press).

MacAskill, William. 2014. *Normative Uncertainty*. DPhil dissertation

(University of Oxford, Department of Philosophy).

MacAskill, William. 2016. Normative uncertainty as a voting problem. *Mind*, 125.500, 967-1004.

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

MacAskill, William, Cotton-Barratt, Owen and Ord, Toby. 2020. Statistical normalization methods in interpersonal and intertheoretic comparisons. *Journal of Philosophy*, 117.2, 61-95.

MacAskill, William and Ord, Toby. 2020. What maximize expected choice-worthiness? *Noûs*, 54.2, 327-53.

Nash, John F. Jr. 1950. The bargaining problem. *Econometrica*, 18.2, 155-62.

Newberry, Toby and Ord, Toby. 2021. The parliamentary approach to moral uncertainty. Future of Humanity Institute, technical report 2021-2.

Nissan-Rozen, Ittay. 2015. Against moral hedging. *Economics and Philosophy*, 31.3, 349-69.

Nozick, Robert. 1974. *Anarchy, State, and Utopia* (New York, NY: Basic Books).

Oddie, Graham. 1994. Moral uncertainty and human embryo experimentation. Pp. 144-61 in K. W. M. Fulford, Grant Gillett, and Janet Martin Soskice (eds), *Medicine and Moral Reasoning* (Cambridge: Cambridge University Press).

Pittard, John and Worsnip, Alex. 2017. Metanormative contextualism

and normative uncertainty. *Mind*, 126.501, 155-93.

Pivato, Marcus. 2022. Review of *Moral Uncertainty*. *Economics and Philosophy*, 38.1, 152-8.

Rechnitzer, Tanja. 2020. Precautionary principles. In *The Internet Encyclopedia of Philosophy*. <URL: <https://iep.utm.edu/pre-caut/>>.

Riedener, Stefan. 2021. *Uncertain Values: An Axiomatic Approach to Axiological Uncertainty* (Berlin: de Gruyter).

Risberg, Olle. 2023. Ethics and the question of what to do. *Journal of Ethics and Social Philosophy*, 25.2, 376-412.

Sepielli, Andrew. 2009. What to do when you don't know what to do. Pp. 5-28 in Russ Shafer-Landau (ed), *Oxford Studies in Metaethics*, volume 4 (Oxford: Oxford University Press).

Sepielli, Andrew. 2010. *'Along an Imperfectly Lighted Path': Practical Rationality and Normative Uncertainty*. PhD dissertation. New Brunswick, NJ: Department of Philosophy, Rutgers University.

Sepielli, Andrew. 2014. What to do when you don't know what to do when you don't know what to do *Noûs*, 48.3, 521-44.

Steele, Katie and Stefánsson, H. Orri. 2020. Decision theory. In Edward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2020). <URL: <https://plato.stanford.edu/archives/win2020/entries/decision-theory/>>.

Sung, Leora. 2023. Supererogation, suberogation, and maximising expected choiceworthiness. *Canadian Journal of Philosophy*, 53.5, 418-32.

Tarsney, Christian J. 2021. Vive la différence? Structural diversity as a

challenge for metanormative theories. *Ethics*, 131.2, 151-82.

Tarsney, Christian J. 2024. Metanormative regress: an escape plan. *Philosophical Studies*, 181.5, 1001-23.

Tenenbaum, Sergio. 2017. Action, deontology, and risk: against the multiplicative model. *Ethics*, 127.3, 674-707.

Thomson, William. 1994. Cooperative models of bargaining. Pp. 1237-84 in Robert Aumann and Sergiu Hart (eds), *Handbook of Game Theory With Economic Applications*, volume 2 (Amsterdam: Elsevier).

Timmerman, Travis and Cohen, Yishai. 2019. Actualism and possibilism in ethics. In Edward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2019).

<URL: <https://plato.stanford.edu/archives/fall2020/entries/actualism-possibilism-ethics/>>.

Vanderschraaf, Peter. 2023. *Bargaining Theory* (Cambridge: Cambridge University Press).

Volij, Oscar and Winter, Eyal. 2002. On risk aversion and bargaining outcomes. *Games and Economic Behavior*, 41.1, 120-40.

Wedgwood, Ralph. 2013. *Akrasia* and uncertainty. *Organon F*, 20.4, 484-506.

Wedgwood, Ralph. 2017. Must rational intentions maximise utility? *Philosophical Explorations*, 20.S2, 73-92.