Chapter 6
Rule Consequentialism and Non-identity

Tim Mulgan

Abstract This paper explores the relationship between rule consequentialism and the non-identity problem. It argues that rule consequentialism accommodates person-affecting intuitions without abandoning Parfit's no difference view. The paper also offers a new model of rule consequentialism—reinterpreting its various features as a series of departures from an act consequentialist ideal each motivated by human finitude and fallibility.

Keywords Future generations · Rule consequentialism · Reproduction · Non-identity · Person-affecting · Parfit · Hooker.

6.1 Introduction

This paper explores a rule consequentialist solution to the non-identity problem. In doing so, I will develop some themes from my recent book *Future People*—and respond to emerging criticisms of that book and of rule consequentialism in general.

My principle aim in *Future People* is to construct a new consequentialist account of the morality of our decisions regarding future people—from individual reproductive choices to global public policy priorities. *Future People* offers the first systematic rule consequentialist account of reproductive ethics, and of the significance of reproductive freedom, and also a new foundation for a liberal theory of intergenerational and international justice.

The present paper has a more limited scope than *Future People*, and also a different emphasis. Its scope is limited in two ways—I focus exclusively on moral theory, and, within moral theory, exclusively on rule consequentialism. One of my subsidiary aims in *Future People* was to motivate a return to the utilitarian tradition in political philosophy, and I regard the discussions of political philosophy and public policy as one of the main features of my book. However, as commentators have focused on the moral side of my project, and as my explorations of political

T. Mulgan (✉)
University of St Andrews, Scotland, UK
e-mail: tpm6@st-andrews.ac.uk

M.A. Roberts, D.T. Wasserman (eds.), *Harming Future Persons*, International Library of Ethics, Law, and the New Medicine 35, DOI 10.1007/978-1-4020-5697-0_6,
© Springer Science+Business Media B.V. 2009

philosophy all rest on a foundation of moral theory, I concentrate here on explaining and defending that foundation.

In *Future People*, I have taken as my primary example the familiar objection that consequentialism is implausible because it makes unreasonable demands. Indeed, one of my aims in writing *Future People* was to bring together the two distinct literatures on obligations to future people and on the demands of morality. In the present paper, however, I focus on the non-identity problem instead. This is partly to fit the theme of the present collection. But I also have a more principled rationale. Largely due to pressure from commentators, I have come to regard my emphasis on the demandingness objection in *Future People* as, at best, misleading. While I still think there are important links between these two problems facing consequentialism, there are also important differences. The most serious problems facing any consequentialist account of future people lie at the intersection between non-identity and demandingness.

In addition to these changes of subject matter, the present paper also seeks to advance beyond *Future People*, by presenting replies to two key objections. The shift from demandingness to non-identity is one such reply. The second is my defence of a contingent morality in Sections 6.6 and 6.7. In several places, my argument in *Future People*, like any exercise in consequentialist ethics, rests on controversial empirical claims. This makes my theory appear vulnerable. Whether we regard this contingency as an objection depends on our views regarding the relationship between moral theory and empirical fact. I shall argue that the most plausible account of that relationship vindicates my approach.

6.2 Two Decisive Intuitions

Contemporary moral theory often begins with moral intuitions—judgements about particular cases or general ideals. The non-identity problem itself is significant because it generates a clash between our moral intuitions and the deliverances of some familiar modes of ethical thinking. The same is true of other puzzles in this area, such as the repugnant conclusion, the mere addition paradox, and the infinite utility puzzle. Intergenerational ethics is especially intuition-based.

I find it helpful to distinguish two kinds of intuitions: *decisive intuitions* (that any acceptable moral theory must accommodate) and *distinguishing intuitions* (that mark distinctive features of different theories). My aim in *Future People* is to develop a theory that accommodates all decisive intuitions, and also makes sense of a range of intuitions that are distinctive of a moderately radical utilitarian outlook.

If we all always agreed in our considered moral judgements, then all our intuitions would be decisive. However, such agreement is not to be found. Sometimes intuitions serve, not to confirm or refute theories, but to distinguish them. There is no definite line between decisive and distinguishing intuitions. No intuition is uncontroversially decisive, if only because there is always a niche in the philosophical marketplace for the first person who rejects it. Partisans of particular moral

theories often present an intuition as decisive, when their opponents would see it as distinctive of that particular theory.

In *Future People*, I begin with two decisive intuitions—two judgements that any acceptable moral theory must respect. These provide test cases for moral theories. In the area of future generations, this test is not trivial, as many familiar moral theories have great difficulty accommodating one or other of these intuitions.

> *The basic wrongness intuition*. It is wrong to gratuitously create a child whose life contains nothing but suffering.
> *The basic liberty intuition.* There is no obligation to have children, nor an obligation not to.

I focus on two simple principles that each have difficulty with one of our basic intuitions. The two principles are as follows.

> *The simple person-affecting principle*. An action can only be wrong if some particular person is worse off than that person would have been if some other action had been performed instead.
> *Simple consequentialism*. The right action in any situation is whatever produces the most valuable state of affairs.

In *Future People*, I use the simple person-affecting principle to illustrate the problems facing non-consequentialist accounts of future morality. These problems owe their prominence to the work of Derek Parfit.[1] Parfit distinguishes two kinds of moral choice. A *same people choice* occurs whenever our actions affect what will happen to people in the future, but not which people will come to exist. If our actions do affect who will come to exist in the future, then we are making a *different people choice*.

Parfit also further distinguishes two kinds of different people choices: *same number* (where our choice affects who exists, but not how many people exist), and *different number* (where we decide how many people ever exist). This second distinction is relevant because simple consequentialism, which seems to cope well in same number choices, faces many difficulties when we turn to different number choices.

Parfit makes three central claims.

1. Different people choices occur very frequently, and in situations where we might not expect them.
2. It is often difficult to tell, in practice, whether we are dealing with a same people choice or a different people choice.
3. Many traditional moral theories cope much better with same people choices than with different people choices. Our moral theories are designed for same people choices, and thus need to be amended to apply to different people choices.

These three claims constitute the *non-identity problem*, so called because, in a different people choice, those who will exist in one possible outcome are not (numerically) identical to those who will exist in an alternative possible outcome.

I must stress that, for the purposes of this paper, I take the non-identity problem to be the *general problem* that arises because we need to adapt same people moral theories to different people choices—and to situations where we are uncertain what type of choice we face. The non-identity problem is not a specific case, a specific intuition, or a specific objection. Nor, as I shall argue, is it a problem only for one theory or class of theories. In particular, non-identity is not only a problem for non-consequentialists.

The non-identity problem is a significant threat to anyone who endorses the simple person-affecting principle, as the latter clearly violates the basic wrongness intuition. If different actions bring different people into existence, then, whatever action we choose, we cannot afterwards locate any particular person who is worse off than he or she would otherwise have been. If we cannot compare existence with non-existence, then we can make no sense of the claim that *x* is worse off than if *x* had never existed. It follows that no simple person-affecting theory can ever condemn any creation choice, however horrific the resulting life.

The major alternative to any person-affecting approach is consequentialism.

Simple consequentialism seems untroubled by the non-identity problem. It easily accommodates the basic wrongness intuition in both different people and same people choices, as it is always wrong to produce less happiness than you might.

Unfortunately, while it does respect the basic wrongness intuition, simple consequentialism clearly violates the basic liberty intuition, as it always obliges us to do *whatever* maximises the good, and thus leaves almost no room for *any* liberty.[2] In any situation, either agents will be obliged to have children (to produce more happy people), or they will be obliged not to have children (because their resources would do more good if devoted to charity). Neither of these obligations is intuitively plausible.[3]

By contrast, the simple person-affecting principle has no difficulty with the basic liberty intuition. This is hardly surprising, as the problem with this principle is that it grants potential reproducers too much liberty, not too little.

Our two simple principles are very crude and over-simplified. Both the person-affecting approach and consequentialism have their defenders, who attempt to accommodate (or explain away) the decisive intuition that is problematic for the simplified version. The non-identity problem began life as an attack on person-affecting views. However, I believe that the situation is now largely reversed, and it is consequentialists who are on the back foot. There are two reasons for this: person-affecting theories thrive, while consequentialism has yet to put its own house in order.

The person-affecting view has many defenders, including many of the contributors to this volume.[4] They argue that such a view can respect the basic wrongness intuition, as we *can* reasonably regard a life not worth living as worse *for that person* than non-existence. Person-affecting theorists also seek to generate stronger obligations regarding future people. One common defence is as follows. The main underlying person-affecting intuition is that an action is only wrong if someone is *wronged*. But a person can be wronged even if it is not the case that they would otherwise have been worse-off. (The classic example is when a person is prevented

from boarding a plane because of his race, and the plane goes on to crash. This person has been wronged—even though he would otherwise have died.) Applying this lesson to the non-identity problem, we can still say that a person has been wronged by an act leading to their creation, even if their life is worth living *and* they would otherwise not have existed at all.

I shall proceed on the assumption that my two basic intuitions are decisive, *and* that some extant person-affecting theories do successfully accommodate them both. Faced with these non-consequentialist alternatives, consequentialists must show *both* that they can accommodate the two basic intuitions, *and* that their theory offers something that even sophisticated person-affecting theories cannot. Before explaining the resources and advantages of my moderate consequentialism, we must ask exactly how—and why—simple consequentialism fails.

6.3 How Simple Consequentialism Fails

Simple consequentialism gets many things right. In addition to respecting the basic wrongness intuition, it also respects several other common intuitions regarding future people. I will examine two examples: gratuitous sub-maximisation and the no difference view.

> *The gratuitously satisficing mother*. Betty has decided to have a child. She could have one in summer or in winter. A child born in winter will not suffer any serious ailments or disabilities, but he or she will have a lower quality of life than a child born in summer. Betty herself is completely indifferent when she has her child. On a whim, Betty decides to have her child in winter.

As the resulting child has a very worthwhile life, it is hard to see how any person-affecting theory could fault Betty's choice. By contrast, simple consequentialism clearly implies that Betty ought to create the child with the better life.

This is a case of blatant moral satisficing, where an agent deliberately produces a sub-optimal outcome on the grounds that it is "good enough," even though she could have produced a significantly better outcome at absolutely no cost to herself. The rationality and morality of satisficing behaviour have been much discussed. I and others have argued elsewhere that blatant satisficing is clearly unjustified in same people choices.[5] Why should we permit it in different people choices? If other things are completely equal, what possible justification is there for such a blatant failure to produce a person with a better life?

This tale generates intuitions that are much harder to avoid for a person-affecting theory than the basic wrongness intuition. On the other hand, these new intuitions are much less forceful. Proponents of the person-affecting approach may simply deny that Betty's choice is wrong. Indeed, they can see its verdict in this case as yet another strike against consequentialism.

I agree that this thought experiment generates no decisive intuitions. However, it does bring out a cluster of intuitions that are problematic for the person-affecting approach. It is at least plausible to believe that there is good reason to opt to create

the more valuable life over the less valuable one; that one ought to do so if other things are equal; and that the source of these reasons lies in the fact that the former option leads to a more valuable outcome—even if that outcome is better for no-one. Not everyone shares these intuitions. For those who do, however, they provide one motivation for exploring alternatives to the person-affecting approach. (I suggest in Section 6.7 that the ultimate fate of these particular intuitions may rest on our ability to reconcile them with the asymmetry intuitions explored in the rest of this section.)

The no difference view is even more controversial. Consider a variant on our previous tale.

> *The two mothers*. Suppose two women (Debbie and Sally) have each decided to have a child. Both must choose between having a child in summer or in winter, where the child born in winter will have a lower quality of life than the child born in summer. On a whim, both decide to have their children in winter. However, due to differences in their respective medical conditions, Debbie faces a different people choice while Sally is making a same people choice.

According to simple consequentialism, there can be no moral difference between these two cases. If Sally's action is wrong, then Debbie's action must be wrong to exactly the same degree. Simple consequentialism implies the following.

> *The no difference view.* If A and B are two situations, and if the only difference between them is that A is a different people choice and B is a same people choice, then there is no moral difference whatsoever between A and B.

Even though they aim to respect the basic wrongness intuition, person-affecting views typically reject the no difference view. I cannot think of a genuinely person-affecting moral theorist who thinks there is *no* difference between same and different people choices.[6] If we embrace the no difference view, then this is a strike in favour of simple consequentialism. However, the no difference view is not universally endorsed. Indeed, the literature contains two extreme responses to these cases. Some hold that there is no difference between the two cases, while others claim that, while Sally's choice may well be wrong, Debbie's cannot be.[7] The first response is most naturally combined with a consequentialist theory, while the second is obviously suited to a person-affecting theory.

I believe there is something to be said for both extremes. My aim in *Future People* is to develop and defend a middle road: while there are good reasons for Debbie to opt for a summer birth, perhaps Sally has additional reasons.

The no difference view follows automatically from a more general feature of simple consequentialism.

> *Impersonalism*. The rightness or wrongness of actions depends entirely upon the value produced, without any regard for how that value is distributed across the lives of human beings.

The impersonalism of consequentialism is also what makes the theory notoriously demanding in famine relief cases. Simple consequentialism requires agents to place their own interests on a par with the interests of others. It leaves no leeway for favouring myself, my nearest and dearest, or my own community. In *Future People*, I treat the failure of simple consequentialism to respect the basic liberty intuition

as a particular *instance* of this excessive demandingness. Melinda Roberts has questioned this diagnosis. She suggests that the real problem for simple consequentialism in regard to future people is not a demandingness problem but instead the aggregative calculation that simple consequentialism is often associated with.[8] I agree that my previous focus on demandingness is, at least, misleading. However, I also think there is some connection between the two problems. The intuitive problems facing simple consequentialism in future morality result from the combined impact of its commitment to demandingness and to the no difference view; and these specific commitments are both instances of simple consequentialism's deeper commitment to impersonalism. The demands of consequentialism are *especially* counterintuitive in different people choices.

Simple consequentialism must endorse the no difference view. If we reject that view, then we must reject simple consequentialism. Why might we reject the no difference view? The main reason is that this view conflicts with a range of intuitive asymmetries, such as the following.

> *The basic asymmetry*. There is no obligation to have children, even if they would be extremely happy. But there is an obligation not to knowingly create people whose lives are not worth living.

This strong asymmetry is a very basic feature of commonsense morality.[9] Imagine a couple who deliberately create a severely disabled child whose life contains absolutely nothing but excruciating agony—simply to explore their own capacity for other-regarding behaviour. Almost no one would find such behaviour morally acceptable. Many people also believe that it is wrong to reproduce if one cannot ensure that one's child's basic needs will be provided for. Yet almost no one thinks that a decision not to reproduce is wrong—at least, not wrong to anything like the same extent.

The intuition behind the basic asymmetry is, in part, an anti-demandingness intuition. Simple consequentialism is wrong to insist that everyone must always promote the good by always creating happy people. But demandingness alone cannot explain the intuitive difference between the two cases. To see this, consider another contrasting pair of cases.

> *Asymmetric demands*. Suppose Mary and Martha are two affluent people in the developed world. They each face a choice between spending their money on themselves and spending it in a way that maximises the good. Mary's alternative is to donate her money to a charity that assists (already existing) disadvantaged people. Martha's alternative is to create a new happy person. Suppose each alternative produces exactly the same total value. Mary and Martha both spend their money on themselves. Has either done anything wrong? And, have they each done something *equally* wrong?

Simple consequentialism must conclude, not only that Mary and Martha each do something wrong, but also that they are *exactly equally* in the wrong. Many people will reject both claims. In particular, there is a strong intuition that Mary's action is open to moral criticism in a way that Martha's is not. Failing to benefit existing people is morally objectionable in a way (and to a degree) that failing to create new happy people is not.

Intuitively, we believe that we are (at least sometimes) morally free to depart from impersonal maximisation, even when that departure involves failing to benefit an existing person. But we also think that our freedom to depart from impersonal maximisation is distinctly *greater* when that departure involves choosing to create no one at all rather than a very happy person; or choosing to create a happy enough person instead of a happier person; or choosing to give a benefit to an existing person rather than creating a new person.

A similar asymmetry applies to other moral distinctions. Common sense regards causing harm as worse than failing to benefit. It draws a *greater* distinction when the contrast is between creating a person whose life is not worth living and failing to create a person whose life is well worth living. The former is clearly forbidden, while the latter is not blameworthy at all.

These asymmetries relate to different number choices, rather than same *number* different *people* choices. So they differ from the examples standardly used in discussions of the no difference view. However, these new asymmetries clearly bring out the underlying problem for simple consequentialism—that it cannot take account of the identity of persons.

6.4 Why Simple Consequentialism Fails

These intuitive failings can be traced (in part) to the fact that (at least in the literature on future generations) simple consequentialism is usually combined with the following account of value.

> *The total view.* The value of a state of affairs is entirely a function of the total well-being it contains, and is unrelated to the distribution of well-being across persons.

One obvious solution is thus, not to reject simple consequentialism itself, but rather to reject the total view. Many consequentialists take this route for independent reasons—largely driven by puzzles in value theory such as the repugnant conclusion, the mere addition paradox, or the infinite utility problem.

Others reject simple consequentialism, adopting a moderate moral theory. I take this second route. I claim that my view has several advantages—or, at least, several distinctive features—when compared to other moderate views. The first is that it is compatible with the total view, and will thus appeal to anyone who wants to retain that view (which has many virtues, and many able defenders[10]), but to combine it with a moderate account of moral obligation.

Other moderate moral theories are compatible with the total view. Most obviously, any theory where moral obligation is independent of the values of states of affairs is consistent with any account of those values. However, my approach is distinctive in retaining from simple consequentialism both the total view and the idea that morality is ultimately all about the promotion of objective value. This distinguishes my approach from those who achieve moderation only by severing or weakening the connection between value and obligation.

A second advantage is that rule consequentialism can be combined with many alternative value theories. Indeed, I argue in *Future People* that most departures from the total view canvassed in the literature would reinforce rule consequentialism in its departures from simple consequentialism.[11] My theory can thus also appeal to those who reject the total view.[12]

A third advantage is that rule consequentialism builds on a departure from simple consequentialism that is already required before we turn our attention to future people. Even though it is more threatening in different people choices, the demandingness objection also arises starkly in same people choices. Consequentialists thus cannot avoid demandingness merely by altering their theory of value—as all salient alternatives coincide in same people choices. They must abandon simple consequentialism. It is thus worth asking whether the solution we devise for same people choices can also do the (related) job in different people choices. In *The Demands of Consequentialism*, I argued that rule consequentialism offers the best solution to the demandingness objection in (most) same people choices. Therefore, in *Future People*, I apply rule consequentialism to our obligations to future people.

A fourth advantage is that, depending on the details, my rule consequentialism may also be able to accommodate some distinctive intuitions, such as the intuition that gratuitous sub-maximisation is wrong.

The final advantage of my approach is that it offers a new account of the relationship between empirical facts and moral rules. This new account enables rule consequentialism to offer a compelling consequentialist justification whenever it either endorses or rejects distinguishing intuitions. We return to this advantage in Sections 6.6 and 6.7.

Because I think consequentialists are on the back foot regarding non-identity, my primary aim is constructive rather than destructive. Instead of seeking to refute rival theories, I concentrate on showing how rule consequentialism respects our two decisive intuitions.

6.5 Rule Consequentialism

Future People defends a form of rule consequentialism. Acts are assessed indirectly, in terms of an ideal code of rules. I use the following general formulation, based on the recent work of Brad Hooker.[13]

> An act is wrong if and only if it is forbidden by the code of rules whose internalisation by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being.

Two features of rule consequentialism play key roles in *Future People*.

1. To assess the costs and benefits of internalising a code of rules, we do not imagine any centrally co-ordinated mass indoctrination. Instead, we assume that moral rules are taught in the normal way—by family, teachers, and the broader culture.

2. We assess the costs of teaching a moral code to a *new* generation. We do not ask what would happen if we tried to teach the new code to a generation of adults who had already internalised a different moral code. This gives rule consequentialism a potential for radical innovation.

Rule consequentialism has been subject to many objections, and much debate, in the recent literature. I address some objections elsewhere, and offer my own solutions.[14] My present focus is purely on rule consequentialism's ability to cope with non-identity intuitions. Can rule consequentialism provide an alternative to moderate person-affecting views?

We begin with our two basic intuitions. A moral code allowing agents to gratuitously create miserable people would not maximise value. Rule consequentialism thus easily respects the basic wrongness intuition. (Rule consequentialism also seems able to accommodate a prohibition on gratuitous sub-maximisation—as a rule telling agents to produce happier people (instead of people who are less happy) will produce better consequences than a rule permitting the creation of less happy people. However, I suggest in Section 6.7 that the relationship between rule consequentialism and gratuitous sub-maximisation is more complex.)

The harder task is to show that rule consequentialism respects the basic liberty intuition. This task lies at the heart of *Future People* and is the focus of most objections. My present project is to broaden the scope of the discussion: to show how rule consequentialism both avoids all the pitfalls caused by the impersonalism of simple consequentialism and accommodates the various personalised asymmetries of common-sense intuition.

6.5.1 Differentiating Rule and Simple Consequentialism

The first step is to differentiate rule consequentialism from simple consequentialism. Among many other failings, simple consequentialism cannot respect the basic liberty intuition. It is thus an unacceptable theory. Rule consequentialism can only be an acceptable theory if it diverges from simple consequentialism. The ideal code of rules cannot be identical to the rule—"Always do whatever produces the best consequences."

To avoid the collapse into simple consequentialism, rule consequentialists seek a middle ground between overly simplistic rules and infinitely complex ones. Many contemporary formulations of rule consequentialism are driven by the need to avoid the collapse into simple consequentialism. I borrow my reply from Hooker, who introduces the distinction between "following a rule" and "accepting a rule" largely for this purpose.[15]

The acceptance of a rule by a population has consequences over and above compliance with that rule. Some people might accept a rule even though they do not always comply with it, while others might comply perfectly with a rule they do not accept. For instance, many people accept, on some level, more demanding principles regarding donations to charity than they can bring themselves to fully comply with,

while social or legal sanctions often produce compliance without genuine acceptance. To accept a rule involves many things other than a disposition to comply with that rule, such as the disposition to encourage others to comply, dispositions to form favourable attitudes toward others who comply, dispositions to feel guilt or shame when one breaks the rule and to condemn and resent others' breaking it, etc.

In *Future People*, I defend a form of rule consequentialism that relies heavily on the requirement that rules be accepted and internalized. This theory incorporates a clear distinction between acceptance and compliance. If the form of rule consequentialism defended in *Future People* is a coherent moral theory, then it does not collapse into simple consequentialism. *If* such a heavy emphasis on internalisation can itself be justified, *then* rule consequentialism is a distinct theory. However, this places even more pressure on my use of internalization, which I defend in Section 6.6.

The differentiation from simple consequentialism is, of course, only the beginning. The crucial question is whether rule consequentialism can use the gap between the two theories to provide an intuitive response to the non-identity problem.

6.5.2 Rule Consequentialism and Reproductive Freedom

Except for one or two brief comments, Hooker himself does not apply his theory to future generations. Indeed, I could find no detailed rule consequentialist account of either individual reproduction or inter-generational justice. One main purpose of *Future People* was to construct such an account. I argue at length that rule consequentialism does support a wide range of commonsense individual freedoms, including reproductive freedom. A crucial starting point is Hooker's observation that the question to which rule consequentialism is the answer is not "what if everyone did that?" but rather "what if everyone felt free to do that?" Hooker himself explicitly, if very briefly, applies this distinction to the morality of reproduction.[16]

> Suppose my nephew tells me he refuses to have children. If everyone refuses to have children, the human species will die out. This would be a disastrous consequence. But it is irrelevant to the morality of my nephew's decision. What is relevant is that everyone's feeling free not to have children will not lead to the extinction of the species. Plenty of people who do not feel obligated to have children nevertheless *want* to—and, if free to do so, will. Thus, there is no need for a moral obligation to have children. Neither is there any need for a general moral obligation to have heterosexual intercourse.

I begin by establishing a strong prima facie case for reproductive freedom. I borrow from J. S. Mill's classical utilitarian defence of liberty, market freedom, and democracy. Given the nature of human beings, things go better overall if people are free to make significant moral decisions for themselves. Arguments against reproductive freedom are then examined and found wanting. *Future People* draws on a range of empirical evidence to argue that reproductive freedom is not a threat to the survival and flourishing of humanity. This leads to my defence of personal liberty and democratic institutions. While not infallible, they promote human happiness and offer the best safeguards for human survival.

Any calculation of the likely results of teaching a code of rules to a new generation involves great uncertainty. This uncertainty may seem a weakness of rule consequentialism—and many philosophers have argued that it is.[17] But, in *Future People*, I argue that uncertainty is really a strength of rule consequentialism. The rules regarding reproductive freedom I develop in Chapter 6 of *Future People* are very general and leave considerable room for judgement in their application. I argue that, given the uncertainty of their future circumstances, it is better to teach the next generation these flexible general rules than to teach them a specific code tailored to the particular dilemmas we expect them to face in the future.

6.5.3 Rule Consequentialism and Person-Affecting Elements

My rule consequentialism has an impersonal foundation—the total view. This distinguishes it from other moderate moral theories. Simple consequentialism, when combined with the same impersonal foundation, yields a morality whose *content* is fully impersonal. To avoid an impersonal content, my rule consequentialism must include a range of obligations to particular people in its moral code. The best code that can be taught to human beings will include obligations to keep promises and to help friends, along with a range of other commonsense moral rules, such as prohibitions on murder and theft. This fit with conventional morality is often presented as a major benefit of rule consequentialism.

Accordingly, while it rejects a person-affecting *foundation* for morality, rule consequentialism need not reject all person-affecting elements within morality. The ideal code may include person-affecting rules and attitudes. Indeed, in *Future People*, I argue that it does include them. Recall that we are asked to imagine a moral code taught in the normal way in the context of a small set of interpersonal relationships. Any moral code is thus learnt via (specific) person-affecting rules. It is then natural to carry these rules (and their accompanying attitudes and moral outlook) over into the rest of our moral lives—even into different people choices.

There is a tension between these person-affecting arguments and the impersonal foundation of rule consequentialism. And there are limits on the content of the ideal code. Any code will include a general disposition to be benevolent, as the benefits of such a disposition are obvious. And *no* code will include the simple person-affecting principle. Someone who has internalized the ideal code will not plant a bomb in a forest that will explode in two centuries—even if they know that, because the act of setting the bomb will alter the identity of all future people, no particular future person will be worse off as a result of this action. We ourselves have learnt a code that (in its application to different people choices) departs from the simple person-affecting principle and produces better results than any code incorporating that principle. If we have learnt a better code than any simple person-affecting code, then no such code can be the best code humans could be taught.

On the other hand, we haven't learnt a code that goes to the other extreme. We have not internalized the no difference view. Furthermore, in *Future People*,

I argue that we *could* not internalize that view.[18] The no difference view requires full impartiality. Partiality—of any kind—is only possible if we attach significance to the numerical identity of persons. Yet humans cannot internalise a fully impartial code, as such a code would be impossibly demanding.

6.6 A Contingent Morality

Having outlined the basis of my rule consequentialist response to the non-identity problem, I now turn to one common objection. My reply to this objection will lead to further elaboration of rule consequentialism.

Several reviewers of *Future People* object to my extensive use of empirical claims in defending rule consequentialism.[19] In particular, they argue that I over-use the device (borrowed from Hooker) of rejecting counterintuitive rules on the basis of controversial empirical claims about what could (or could not) be taught to a population of human beings. For instance, to establish that rule consequentialism will not require agents to sacrifice all their own interests for those of future people, I claim that any population of humans would regard such a rule as unreasonably demanding—and thus that it cannot be successfully taught. Internalization is thus central to both my strategy for differentiating rule consequentialism from simple consequentialism and my attempt to justify reproductive freedom.

I aim to show that the rule consequentialist reliance on internalization costs is not under-motivated, and that rule consequentialism is not inappropriately reliant on empirical accidents. I also argue that, far from being a weakness, my reliance on empirical facts points to another advantage of my account—its ability to offer a plausible unifying story of the role of both empirical information and philosophical debate in the moral life of human beings.

I must begin by conceding that rule consequentialism's intuitive appeal *is* entirely contingent. Even in regard to the most decisive intuitions, rule consequentialism only gives the right answers because of (contingent) empirical factors. The reason for this is simple. Rule consequentialism offers a series of reasons to depart from simple consequentialism. Each of these reasons is built, ultimately, on a claim that is contingent. Things could have been very different. If they had been different, then simple consequentialism would have been the best moral code. There are thus possible worlds where rule consequentialism collapses into simple consequentialism. As simple consequentialism violates decisive intuitions such as the basic liberty intuition, it follows that it is only contingently true that rule consequentialism respects decisive intuitions. And there may be alternative moral theories that do not rest on such contingencies. (Consider a libertarian morality, where the demands of morality depend only on the agent's own voluntarily assumed obligations.) If it counts against a moral theory that it answers moral questions with contingent facts, then rule consequentialism is at a significant comparative disadvantage.

Even rule consequentialism's respect for the basic wrongness intuition is contingent, as it only endorses that intuition because the consequences of teaching a code requiring agents to take account of the interests of future people (even in different

people choices) are better than the consequences of teaching any code permitting disregard of future people. This comparative claim may seem obviously true. But it is significant to note that, however obvious, it still rests on contingent empirical features of human beings—not merely on logical features of rules, or on impersonal values. We can imagine creatures so deeply ingrained with a lack of interest in future people—or so wedded to the simple person-affecting principle—that any attempt to teach them *any* obligations in different people choices would be counter-productive. Rule consequentialism only endorses the basic wrongness intuition because we are not such creatures. Should this "contingency" worry us? I suggest that it should not.

6.6.1 Defending Internalisation

This brings us to my defence of rule consequentialism. Morality is for *creatures like us*. The contingent facts I appeal to in *Future People* are deep facts that make us what we are. It is a strength of rule consequentialism—not a weakness—that its moral verdicts apply only to creatures like us and only in situations (broadly) similar to our own.

Most will agree that morality should appeal to some contingent facts. But my argument doesn't just appeal to some facts. Instead, it rests very heavily on one particular set of facts—those relating to the costs of teaching rules to human communities. Why are *those facts* so important to morality?

To provide a more solid defence than I offered in *Future People*, I now seek to explain why the focus on internalization in particular is a response to a plausible rule consequentialist story about the role of morality—and not an ad hoc device introduced merely to render rule consequentialism more user-friendly.

Rule consequentialism regards morality as a code of rules to enable a community of human beings to live together in a way that promotes human well-being and human flourishing. It is important to note that this *not* an evolutionary, descriptive, or semantic claim—but a *normative* claim. I am not saying any of the following: "This is why morality evolved," "This is what 'morality' means," "This is what morality (empirically) is." Rather, *Future People* develops a rule consequentialist suggestion as to how we might usefully answer the question: "Why is morality important to us?"

Rule consequentialism's basic question is this: What would happen if a code of rules (R) were to become the moral code for a community of human beings—*by the standard natural process*? For (perhaps deceptive) ease of presentation, in *Future People* I usually put this question in first-person plural terms for the present generation. (What would happen if *we* tried to teach R to the next generation?) But the focus is meant to be on the *teachability of the code*, not on *our ability to teach*. This interpretation is a logical extension of our motivation for abandoning simple consequentialism in the first place. If we are sympathetic to rule consequentialism at all, then it makes little sense to ask what would happen if R became a moral code for human beings *as if by magic*. Why would anyone be interested in *that*

question? Either we interpret the utilitarian tradition at an abstract level or we seek to apply it to the situation of real human beings. The former route leads to simple consequentialism, the latter to a form of rule consequentialism that asks what would happen if rules were taught to humans in the usual way.

The costs of teaching a code reflect that code's degree of fit with human beings and their situation. This provides a useful measure of the code's suitability as a moral code *for humans*. Rule consequentialism is right to place weight on such facts, and to use them to differentiate itself from simple consequentialism.

6.6.2 Freedom and Person-Affectingness Revisited

Having sketched a general defence of internalisation, we turn now to reconsider the two key features of my response to the non-identity problem: freedom and person-affectingness.

I begin with the rule consequentialist defence of freedom from Chapter 6 of *Future People*. This argument is very clearly not a priori, as it explicitly cites empirical studies made prominent by the work of Amartya Sen.[20] My central claim in *Future People* is that, as a matter of fact, given the kinds of creatures human beings turn out to be, things go better overall (in terms of human well-being broadly construed) if people are left to make major life choices (especially reproductive choices) for themselves rather than having those choices made for them. Any such argument is, of course, heavily dependent on (empirical) claims as to *how* people will exercise this freedom. The argument for reproductive freedom only goes through if we can be reasonably confident that people will not respond to such freedom in a way that leads to underpopulation or overpopulation.

Freedom is morally appropriate *for us*. But we can easily imagine creatures for whom it is not. There are possible creatures in other possible worlds whose well-being is maximized by coercion rather than choice. Insofar as it says anything about those creatures (and there is, by the way, no reason why it *should* say anything), rule consequentialism must say that, for them, the appropriate moral code will sanction (and perhaps require) widespread coercion.

All rule consequentialist arguments for moral freedom—of any kind—share this contingency. Freedom has obvious costs from an impersonal consequentialist point of view—as it leads to sub-optimal decision-making in some circumstances. (If all agents are allowed to refrain from maximising the good, then some will so refrain.) Therefore, to be included in the rule consequentialist ideal code, any freedom must have compensating benefits. These benefits arise because of contingent features of our nature—including, perhaps, the fact that we are creatures for whom freedom is an independently valuable component of well-being.

I argued above that, in addition to supporting reproductive freedom, rule consequentialism avoids the no difference view. As ever, my argument was not a priori. From a consequentialist point of view, more impersonal rules would offer the best fit with the total view. A community of rational agents who perfectly follow a no difference code would produce better results than one following a person-affecting

code. To avoid the no difference view, rule consequentialism must show that this view cannot be effectively internalized.

The empirical case here is similar to that offered by rule consequentialism regarding demandingness—only stronger. The no difference view is extremely impersonal. No code for creatures (remotely) like us can be nearly so impersonal. Perhaps we can imagine perfect utilitarian calculating machines who would be best suited to a simple consequentialist code that accords absolutely no moral significance to the numerical identity of persons. But these imaginary agents are unlike us in very morally relevant ways.

Does all this contingency undermine rule consequentialism? I think not. The "contingency" underlying my defence of freedom in *Future People* is not coincidence or accident. It reflects general and important features of human nature and the human situation. Why shouldn't the content of an ethic for humans depend on such features? Indeed, on what else could it depend? On pure reason? On disembodied rationality? Rule consequentialism takes its inspiration from J. S. Mill and other classical utilitarians. Human nature is something we discover empirically, not something we intuit a priori. If future empirical studies overturned our views about human nature, then we would (and, rule consequentialism argues, we *should*) amend our moral views.

6.7 Rule Consequentialism and Moral Philosophy

If we allow it to appeal to contingent facts, then rule consequentialism can respect all decisive intuitions—those that, even on reflection, we cannot imagine giving up. Rule consequentialism reinterprets decisive intuitions as those we cannot imagine fitting together with *any* moral code that could be effectively internalised by a community of human beings. We simply don't think that human beings *could* live *well* like *that*.

We must note that the notion of "effective internalisation" is itself cashed out in consequentialist terms. Rule consequentialism doesn't deny that an individual human being or a community might train themselves (or be trained by some outside agency) to believe in simple consequentialism, or to have no regard for distant future people. What it denies is that these rules could be part of a moral code that maximises human well-being over the long-term.

Recall our distinction between decisive intuitions and distinguishing ones. Distinguishing intuitions are the sites of controversy in moral philosophy. Rule consequentialism offers an account of that controversy. A distinguishing intuition is one where we are not sure if it fits with the best ideal code or not. And, says rule consequentialism, we decide whether to embrace a controversial intuition *by* asking how well it fits with such a code.

Many people find this last claim implausible. Surely the way we decide between controversial intuitions bares little resemblance to rule consequentialist inquiry? I now seek to dissolve this objection, by bringing moral philosophy itself within the rule consequentialist framework.

Rule consequentialists do not expect to discover the ideal code in its entirety in one go.[21] Instead, we discover that a particular rule—or a rule of some general type—is in the code. Given the way that human beings happen to be, a code that permits favouring self and nearest and dearest, forbids murder, obliges promise-keeping, and promotes beneficence will produce better results than one that does not. We know the ideal code includes these elements—even if we cannot hope to describe that code its all its details.

Decisive intuitions provide constraints—fixed points in the moral psychology of someone who has internalised the ideal code. Once these parameters are set, we explore controversial intuitions by asking how someone who had internalised these (decisive) rules would (most naturally) respond to other situations.

At this point in its inquiry, rule consequentialism welcomes a wide variety of (often inconclusive) empirical evidence, which can enlighten us on the limits and flexibility of human moral codes. If human beings have been effectively taught a code with rule R, then we at least know that codes with rule R can be taught to human beings. Rule consequentialism thus offers a sound consequentialist argument for borrowing moral rules from other cultures, so long as those rules work better than our current commonsense morality.

Another source of evidence within rule consequentialism is the progress of moral philosophy itself. For instance, suppose philosopher P develops an intuitively plausible and coherent moral theory, which links decisive intuitions together using plausible moral ideals. P's achievement then itself constitutes prima facie evidence that such a code makes sense as a moral view of the world—and thus *could* be a moral code for human beings. If our worry about code R is whether R can be (efficiently) internalized, then a coherent account of an intuitively plausible version of R helps alleviate that worry.

Consider a concrete example. Should we extend our prohibition on gratuitous sub-maximisation from same people choices to different people choices? For rule consequentialism, this is the question whether a person who had internalised the general norms of the ideal code would find it natural to bring certain particular cases under her person-affecting dispositions or under her general disposition to promote human well-being—or (somehow) bring such cases under both dispositions. The discovery that a certain pattern of thought makes best sense of our own moral intuitions may help us to decide how this idealised agent would react.

Throughout this paper, I have equivocated as to whether or not rule consequentialism prohibits or endorses gratuitous sub-maximisation in different people choices. But this is because I am unsure which of these attitudes best fits both with the basic wrongness intuition and with our decisive intuitions about same people choices. Considerable further exploration is necessary before we can settle this question. I do claim, however, that both consequentialists and non-consequentialists should be able to agree that rule consequentialism focuses attention on the right question here. Whatever its outcome, the rule consequentialist process promises a justification for one or other distinguishing intuition.

Rule consequentialism can thus borrow from person-affecting moral theories, as these moral theories provide evidence of the internalizability of moral codes.

However, rule consequentialism does not thereby *become* a person-affecting *theory*, as its foundation remains resolutely (and impersonally) consequentialist. Instead, in this contested region between the end-points marked by decisive intuitions, rule consequentialism offers a new way to organize *all useful moral input*.

Suppose we discovered a community of human beings who were clearly flourishing better than ourselves and whose moral code differed from our own. (Perhaps their moral philosophers have taught them to conceptualise Parfit's puzzle cases in a way that we cannot yet imagine, enabling their community to flourish across the generations in a way that ours does not.) Rule consequentialism's central claim is that we would conclude that their moral code was superior—that they had stumbled upon a better way for human beings to live. We would then adopt their moral views in controversial cases—or at least attempt to move our own views in their direction. An intuition is decisive when we cannot imagine encountering such a community. It is distinguishing (or controversial) when we can. (If follows, of course, that the judgement that a certain intuition *is* decisive can only ever be provisional. The fact that we cannot imagine encountering a more flourishing community who lack that intuition, doesn't prove that we won't.)

Such encounters need not be the stuff of exotic anthropology or bizarre science fiction. Judged in consequentialist terms, our present moral code is superior to the codes of earlier generations in many ways. There is no reason to expect our own generation to mark the end of moral progress. We may reasonably hope that future people will have better moral beliefs than ourselves. Rule consequentialism offers an account of what this claim means. It also suggests that, if we can discover what those superior future beliefs might be, we should adopt them for ourselves.

Acknowledgements For helpful comments on earlier drafts of this paper, I am grateful to the editors of this volume. In this paper, I draw freely on Mulgan (2006). For permission to do so, I am grateful to Oxford University Press. I presented an earlier version of my ideas on rule consequentialism and contingency at an international conference on "Utilitarianism: An Ethics of Experience?" held at the University of Rome in June 2007. For the invitation to attend the conference, for their warm hospitality, and for comments on my paper, I am very grateful to Eugenio Lecaldano, Gianfranco Pellegrino, and Francesco Orsi. For comments on the paper, I am grateful also to Brad Hooker and John Skorupski. I am also grateful to Rahul Kumar, Francesco Orsi, Gianfranco Pellegrino, Melinda Roberts, and Rivka Weinberg for their published comments on Mulgan (2006), which stimulated me to write this paper.

Notes

1. Parfit (1984), pp. 351–441.
2. Simple consequentialism can give us some liberty in cases where two or more outcomes are tied for being "the best" in terms of aggregate wellbeing.
3. In *Future People*, I argue that, in the actual world, simple consequentialism is more likely to oblige affluent people in developed countries *not* to reproduce, as they could invariably do more good by giving their money away. Mulgan (2006), pp. 16–20.
4. For discussions of the person-affecting approach, see, for instance, Feinberg (1986); Heyd (1992); Kumar (2003); McMahan (1998); Roberts (1998, 2002, 2003); Temkin (1993); Woodward (1986).

5. For discussion and references, see Mulgan (2001), pp. 127–44.
6. Since the view Nils Holtug suggests in his contribution to this collection would not count as "person-affecting" in my sense of that term (identity, for him, isn't critical to a person-affecting assessment of wrongdoing), he isn't a counterexample to my claim.
7. Parfit defends the no difference view. See Parfit (1984), pp. 366–71. The opposite view is adopted by Heyd (1992) and is implicit in many defences of the person-affecting approach.
8. Roberts (2007), p. 775.
9. We should note that the asymmetry is not uncontroversial, as demonstrated by the papers in this volume by Persson and McMahan.
10. See, for instance, Broome (2004).
11. Mulgan (2006), pp. 142–46.
12. However, there are limits to the flexibility of my account. In particular, I do not see how *rule* consequentialism—which evaluates rules collectively—can be combined with a relativised or person-affecting value theory. Rule consequentialism is thus a rival for the views of Partha Dasgupta and Melinda Roberts. Dasgupta (1993, 1994); Roberts (1998, 2002, 2003).
13. The following exposition of rule consequentialism draws freely on Mulgan (2006), pp. 130–60, which in turn is based on Hooker (2000).
14. For discussion and references, see Mulgan (2001), pp. 53–103; and Mulgan (2006), pp. 130–60.
15. Hooker (2000), pp. 75–80; Mulgan (2006), pp. 138–40. The original "collapse" objection is due to Lyons (1965).
16. Hooker (2000), p. 177.
17. See, for instance, Griffin (1996), pp. 103–7. For further references and discussion, see Mulgan (2006), pp. 150–52 and 244–53.
18. Mulgan (2006), pp. 154–59.
19. See especially Kumar (2007); and Roberts (2007). See also Orsi (2007); Pellegrino (2007); and Weinberg (2006). I reply to some of their concerns in Mulgan (2007a) and (2007b).
20. See, especially, Sen (1999), pp. 204–26.
21. Mulgan (2006), pp. 150–52.

References

Broome, J. 2004. *Weighing lives*. Oxford: Clarendon Press.

Dasgupta, P. 1994. Savings and fertility: Ethical issues. *Philosophy and Public Affairs* 23: 99–127.

Dasgupta, P. 1993. *An inquiry into well-being and destitution*. Oxford: Clarendon Press.

Feinberg, J. 1986. Wrongful life and the counterfactual element in harming. *Social Policy and Philosophy* 4: 145–178.

Griffin, J. 1996. *Value judgement*. Oxford: Clarendon Press.

Heyd, D. 1992. *Genethics: Moral issues in the creation of people*. University of California Press.

Hooker, B. 2000. *Ideal code, real world: A rule-consequentialist theory of morality*. Oxford: Clarendon Press.

Kumar, R. 2003. Who can be wronged? *Philosophy and Public Affairs* 31: 99–118.

Kumar, R. 2007. Review of T. Mulgan, *Future people*. *Philosophical Quarterly* 57: 679–685.

Lyons, D. 1965. *The forms and limits of utilitarianism*. Oxford: Clarendon Press.

McMahan, J. 1998. Wrongful life: paradoxes in the morality of causing people to exist. In *Rational commitment and social justice: Essays for Gregory Kavka*, eds. J. Coleman and C. Morris. Cambridge: Cambridge University Press.

Mulgan, T. 2001. *The demands of consequentialism*. Oxford: Clarendon Press.

Mulgan, T. 2006. *Future people*. Oxford: Clarendon Press.

Mulgan, T. 2007a. *Future people*: una presentazione [Precis of *Future People*]. *Filosofia e Questioni Pubbliche* 12: 135–140.

Mulgan, T. 2007b. Riposti ai commenti [Replies to Commentators]. *Filosofia e Questioni Pubbliche*. 12: 163–185.

Orsi, F. 2007. Dividere la moralita e connettere ragioni e valori. *Filosofia e Questioni Pubbliche* 12: 141–150.
Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
Pellegrino, G. 2007. Liberta riproduttiva e obblighi generali nei confronti delle generazioni future. *Filosofia e Questioni Pubbliche* 12: 151–163.
Roberts, M. 1998. *Child versus childmaker: Future persons and present duties in ethics and the law*. Lanham, Maryland: Rowman & Littlefield.
Roberts, M. 2002. A new way of doing the best we can: Person-based consequentialism and the equality problem. *Ethics* 112: 315–350.
Roberts, M. 2003. Is the person-affecting intuition paradoxical? *Theory and Decision* 55: 1–44.
Roberts, M. 2007. Review of T. Mulgan, *Future People*. *Mind* 116: 770–775.
Sen, A. 1999. *Development as freedom*. Oxford: Clarendon Press.
Temkin, L. 1993. *Inequality*. Oxford: Clarendon Press.
Weinberg, R. 2006. Review of T. Mulgan, *Future people*, *Notre Dame Philosophical Reviews* 2 (ndpr.nd.edu).
Woodward, J. 1986. The non-identity problem. *Ethics* 96: 804–831.