

Act Contractualism

Léa Bourguignon

Thesis submitted as part of the MPhil Stud in Philosophical Studies
UCL (University College London)

I, Léa Bourguignon confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract:

Consider three moral views that have received significant attention in the philosophical literature:

Act Consequentialism: an act is permissible if and only if its performance makes things go best (that is, if and only if it brings about the best state of affairs).

Rule Consequentialism: an act is permissible if and only if it conforms to the best set of rules, where the best set of rules is the set such that things would go better if everyone complied with this set (or accepted this set) than if everyone complied with any alternative set (or accepted any alternative set).

Rule Contractualism: an act is permissible if and only if it is allowed by a set of rules the universal acceptance of which is appropriately justifiable to everyone.

The first two of these views are well-known. The third is a plausible reading of the theory defended by Tim Scanlon. All three are plausible and have enjoyed a great deal of discussion in the literature. Much of this discussion has in part aimed at pushing or addressing various challenges in order to weigh these ethical views against each other.

There is another view, however, which is

Act Contractualism: an act is permissible if and only if it is appropriately justifiable to everyone.

The aim of my project is to carve out a conceptual space for this ethical view, which has so far received almost no explicit attention in the literature. I explain more fully the motivation for considering Act Contractualism by looking at the problems faced by the three other theories – Act Consequentialism, Rule Consequentialism and Rule Contractualism. I then assess the plausibility of Act Contractualism and consider its merits in comparison to Rule Contractualism and Act Consequentialism.

Impact Statement:

By the end of this dissertation, I hope to carve out conceptual space for Act Contractualism, and to show that it is a serious contender to the three more familiar views I have considered. This will open the door to future research into what the most plausible version of Act Contractualism would look like. This project will also make a novel contribution to the literature in moral theory by illustrating the way in which Contractualism can be seen as a framework capable of accommodating many first-order theories. As such, it contributes to setting the scene for a “contractualising project”.

Table of contents

Introduction	5
Chapter 1: Act Consequentialism	12
1. The structure of Act Consequentialism	13
1-1. Axiology.....	13
1-2. Ignorance and false beliefs.....	15
1-3. Justified and unjustified beliefs.....	17
1-4. Uncertainty	19
1-5. Final formulation.....	20
2. Objection 1: Williams on Integrity	22
2-1. Doing and allowing.....	23
2-2. Moral self-conception	25
2-3. Over-demandingness	27
2-4. Special relationships.....	29
2-5. Summary	31
3. Objection 2: Rights and Duties	32
4. Objection 3: Collective Impact	35
5. Objection 4: Aggregation	36
Conclusion	39
Chapter 2: Rule Consequentialism	41
1- Objections to AC	43
1-1. Over-demandingness	44
1-2. Special relationships.....	45
1-3. Rights and duties.....	46
1-4. Collective impact.....	46
2- Does RC address these objections?	48
2-1. A consequentialist rationale behind the first three objections.....	48
2-2. Over-demandingness	50
2-3. Special relationships.....	51
2-4. Rights and duties.....	52
2-5. Collective impact.....	53
3- The Collapse Objection and the Reflective Equilibrium Argument	54
3-1. The objection	54
3-2. Response: Compliance and Acceptance.....	57
3-3. The Reflective Equilibrium Argument	59
4- The Ideal World Objection	60
4-1. The objection	60
4-2. Response: General Acceptance.....	61
4-3. Response: Avoid Disaster	64
5- The Challenge Restated	67
6- A Rule-consequentialist rationale?	71
6-1. Impartial Defensibility	71
6-2. Following RC in the real world is not impartially defensible.....	72
6-3. Another account of Impartial Defensibility?	73
Conclusion	75
Chapter 3: Rule Contractualism	77

1. Rule Contractualism.....	78
1-1. Rule Contractualism in the literature	78
1-2. The structure of Rule Contractualism	81
1-3. Interpretations of the requirement of justifiability.....	82
1-4. Theory of reasons or rationality.....	83
2. Rule Contractualism and RC.....	86
2-1. Similarities: Rule Contractualism, RC and the objections to AC.....	86
2-2. Differences: the contractualist account of morality	89
3. Does Rule Contractualism avoid the IWO?	91
3-1. The objection	91
3-2. Mitigating the objection: flexible rules	93
3-3. Collapse?	95
4. Rule Contractualism and Act Contractualism.....	98
4-1. Reason to resist the collapse: avoid the objections to AC	99
4-2. Reason to accept the collapse: justifiability in the real world.....	100
5. Act Contractualism and AC.....	103
Conclusion.....	106
<i>Bibliography</i>	<i>107</i>

Introduction

Consider three moral views that have received significant attention in the philosophical literature:

Act Consequentialism (AC): an act is permissible if and only if its performance makes things go best (that is, if and only if it brings about the best state of affairs).

Rule Consequentialism (RC): an act is permissible if and only if it conforms to the best set of rules, where the best set of rules is the set such that things would go better if everyone complied with this set (or accepted this set) than if everyone complied with any alternative set (or accepted any alternative set).¹

Rule Contractualism: an act is permissible if and only if it is allowed by a set of rules the universal acceptance of which is appropriately justifiable to everyone.

The first two of these views are well-known. The third is a plausible reading of the theory defended by Tim Scanlon.² All three are plausible and have enjoyed a great deal of discussion in the literature. This discussion has in part aimed at pushing or addressing various challenges in order to weigh these ethical views against each other.

In a nutshell, the existing philosophical discussion of these views looks like this: Act Consequentialism (hereafter AC) is intuitively appealing and has both a plausible rationale

¹ The difference between compliance and acceptance versions of this view will be discussed in Chapter 2.

² Scanlon (1998). I will motivate in chapter 3 the claim that this view, and those defended by Southwood (2010) and Parfit (2011), are versions of Rule Contractualism.

and many plausible implications.³ However, it faces some powerful objections.⁴ Rule Consequentialism (hereafter RC) and Rule Contractualism also have plausible rationales and many plausible implications, and they avoid some of the main objections to AC.⁵ However, it is increasingly recognised that these two views instead face another serious problem, known as the Ideal World Objection.⁶ Since this objection relates to a fundamental structural feature of RC and Rule Contractualism, it seems inescapable. This has led some moral theorists to retreat to AC and focus their efforts towards addressing or mitigating the objections raised against it, which appear less inescapable.⁷

There may be another option, however. Rather than retreating to AC, we could retreat to

Act Contractualism: an act is permissible if and only if it is appropriately justifiable to everyone.

The aim of my project is to evaluate this alternative option, which has so far received almost no explicit attention in the literature. Indeed, Scanlon even claims that drawing a distinction between Act Contractualism and Rule Contractualism is “misconceived”, because there can be no question “about whether justification should involve principles at all”.⁸

³ Bentham (1823), Mill (1871/1969), Singer (1993) and Portmore (2011).

⁴ Williams (1973), McCloskey (1965) and Nefsy (2012).

⁵ Hooker (2002), Scanlon (1998).

⁶ Podgorski (2018), Parfit (2011), pp.312-320.

⁷ Railton (1984); Kagan (2011).

⁸ This is because “to justify an action is to offer reasons supporting it”, and to do this, Scanlon argues, amounts to defending a principle, “namely one claiming that such reasons are sufficient grounds for so acting under the prevailing conditions”. See Scanlon (1998), pp.197-202.

The only exception to the silence on Act Contractualism seems to be an article by Hanoch Sheinman.⁹ Sheinman argues that Act Contractualism and Rule Contractualism have different foundational commitments and are not extensionally equivalent. He suggests that Rule Contractualism owes its initial plausibility to the idea that we should act in ways no one can reasonably object to, and that this idea is better captured by Act Contractualism.

However, Sheinman does not try to evaluate the overall plausibility of Act Contractualism nor compares it to AC. Therefore, the aim of my project is twofold. First, I want to explain more fully the motivations for considering Act Contractualism, by discussing the problems faced by AC, RC, and Rule Contractualism. Second, I want to assess the plausibility of Act Contractualism and consider its merits, especially in comparison to AC.

In **Chapter 1**, I will begin by explaining in more detail what AC is and why we might find it plausible, before presenting five important objections to this view. AC is accused of being incompatible with self-concern and personal commitments, which makes it overly demanding (first) and incompatible with special relationships (second).¹⁰ Third, it is said to neglect rights and duties, by implying that we should lie, steal and harm other people whenever doing so would make things go best.¹¹ Fourth, it is accused of not being able to account for the obligatoriness of taking part in beneficial collective action.¹² Fifth, it is criticised for failing to

⁹ Sheinman (2011). Suikkanen (2020, pp.34-35) mentions the possibility of altering the option set element of Scanlonian Contractualism to compare acts instead of principles, but he does not elaborate on this possibility. Beyond that, as Southwood (2010, p.201 fn.33) puts it, “single-level contractualism is virtually unheard of”.

¹⁰ Williams (1973).

¹¹ McCloskey (1965).

¹² Parfit (1986), pp.75-82.

draw a distinction between one person suffering immensely and a large number of people being mildly indisposed.¹³

In **Chapter 2**, we will see that RC looks able to avoid at least some of these objections. This is because, instead of evaluating acts on the basis of their consequences, RC evaluates acts on the basis of the consequences of the universal adoption of a rule that permits such acts, which leads to different results than AC. Therefore, with respect to the objections to AC, RC seems to fare better than AC and has the benefit of tracking common intuitions more accurately.¹⁴

However, I will also show that RC is vulnerable to another set of interrelated objections. According to the first one, the Collapse Objection, RC is extensionally equivalent to AC and runs into all the same problems. Rule consequentialists address this objection by adopting a Universal Acceptance version of their theory, rather than a Universal Compliance version. However, I will argue that resisting the collapse into AC makes RC vulnerable to another criticism, the Ideal World Objection (hereafter IWO).

According to the original version of this objection, evaluating acts in the real world on the basis of the consequences they would have in an ideal world where everyone accepts rules that permit those acts has absurd implications. Rule consequentialists have amended their theory to address the IWO by making its implications less counter-intuitive. I will grant that they do so successfully, but I will restate the IWO in light of these amendments.

Rather than questioning the plausibility of RC's implications, the challenge now concerns RC's explanatory potential. I will argue that the amendments made to RC threaten to expose

¹³ This is often called the separateness of persons objection. See Brink (1993), Rawls (1999) pp.23-27 and Nagel (1970) pp. 138-142.

¹⁴ Whenever I mention intuitions, I mean reflectively stable, carefully considered and widely shared moral intuitions.

it as a theory crafted ad hoc around common intuitions, which makes the need for a unifying rationale pressing. After considering three candidate rationales, we will see that none of them accounts for RC's prescriptions. Therefore, rule consequentialists must accept that their theory is either implausible, or non-explanatory. Given the strength of the IWO (as I have reformulated it), I conclude that we have good reason to accept the collapse into AC, despite the problems this view faces.

In **Chapter 3**, I will introduce a third view – Rule Contractualism – and ask if it fares better than RC in addressing the objection to AC while avoiding the IWO. We will see that Rule Contractualism, like RC, avoids the problems associated with AC by looking at the cumulative consequences of the general performance of certain acts. The difference is that, instead of rules the universal acceptance of which would make things go best, Rule Contractualism prescribes following rules the universal acceptance of which is appropriately justifiable.

The contractualist commitment to appropriate justifiability rather than to making things go best is a key reason why some favour Rule Contractualism over RC. It has been argued that, by trying to act in ways that are appropriately justifiable, one acknowledges them as rational and reasonable equals to whom justification is due, which promotes valuable relationships of mutual recognition.¹⁵ However, despite this important difference between these views, they have similar implications, and a similar structure. Therefore, Rule Contractualism is vulnerable to the exact same line of objection as RC: it either collapses into an act-based version of the theory (the Collapse Objection) or it has absurd implications (the IWO).

Rule contractualists have ways to make the implications of their theory less implausible. However, I will contend that the first horn of this dilemma – accepting the collapse into an

¹⁵ Scanlon (1998), pp.158-168.

act-based version of Rule Contractualism – is a viable and underexplored option. I will thus present Act Contractualism in contrast to Rule Contractualism. I will argue that those who, like me, are attracted to the contractualist rationale, should in fact prefer Act Contractualism to Rule Contractualism. What the IWO shows is that rules the universal acceptance of which is appropriately justifiable can prescribe acting in ways which are not appropriately justifiable in the real world. So, someone who cares about real-world justifiability should evaluate acts directly, without relying on a process of universalisation. This means accepting an act-based theory.

Finally, I will briefly contrast Act Contractualism and AC. Since these views have different foundations, Act Contractualism would not be redundant even if it turned out to have the exact same implications as AC. This is a good reason to take both views seriously and consider them separately. However, I will show that Act Contractualism and AC are vulnerable to different objections regarding the ethics of future people. While Act Contractualism captures what is known as the Procreation Asymmetry, Act Consequentialism does not; and while Act Contractualism faces what is known as the Non-Identity Problem, Act Consequentialism does not.¹⁶ We thus have good reason to think that Act Contractualism and AC do have different implications, which gives us additional grounds to weigh them against each other.

By the end of this dissertation, I hope to carve out conceptual space for Act Contractualism, and to show that it is a serious contender to the three more familiar views I have considered. This will open the door to future research into what the most plausible

¹⁶ For discussion of the Procreation Asymmetry, see Roberts (2011). For discussion of the Non-Identity Problem, see Boonin (2014).

version of Act Contractualism would look like, and into whether it would be able to address at least some of the main objections raised against AC.

Chapter 1

Act Consequentialism

My aim in this dissertation is to carve out conceptual space for a hitherto neglected moral theory, Act Contractualism. I do this by looking at the merits and limitations of three much more familiar moral theories: Act Consequentialism, Rule Consequentialism and Rule Contractualism. In this first chapter, I begin by introducing and discussing the most well-known of the latter three views – Act Consequentialism (hereafter AC). This will motivate a discussion of Rule Consequentialism in chapter 2, and of Rule Contractualism in chapter 3, since these views have been formulated as an answer to some of the shortcomings of AC. It will also provide us with a benchmark against which to briefly test and compare Act Contractualism at the end of chapter 3.

I start in section 1 by presenting AC, outlining its structure and giving reasons to adopt this theory. The remaining sections are dedicated to what I take to be the main objections to AC. In section 2, I critically reconstruct Bernard Williams' integrity objection, which concerns AC's disregard for the identity of the agent that brings about a given state of affairs.¹⁷ In section 3, I argue that AC neglects moral rights and duties. In section 4, I present the problem of collective impact. In section 5, I offer one last objection against AC, that from aggregation. I conclude that these objections give us good reason to look for a theory that can do better than AC.

¹⁷ Williams (1973).

1. The structure of Act Consequentialism

1-1. Axiology

According to

Act Consequentialism (AC): an act is permissible if and only if it makes things go best (that is, if and only if it brings about the best state of affairs).

As it stands, this definition does not say much. One must specify what it would mean for things to go “best”. This is why AC relies on a theory of value – also called an axiology – which determines what is to be maximised. For instance, according to one version of AC, Utilitarianism, the only thing of value is wellbeing. According to other versions of AC, other things have intrinsic value, such as equality, natural beauty, or human relationships.

AC can also value aspects of actions which do not, strictly speaking, pertain to their consequences. It can attribute intrinsic value or disvalue to other features of an action, or even to certain types of action. For instance, act consequentialists may want to attribute disvalue to the act of lying. In this case, as Jussi Suikkanen puts it, they might consider that “one outcome of telling a lie will be that a lie has been told”.¹⁸

By tweaking their axiology, act consequentialists can thus construct a wide variety of theories, which can have very different, often conflicting implications, some more plausible than others. This allows them to address some of the objections raised against (versions of) AC. Take, for example, Utilitarianism. Since the only thing that matters to our moral evaluation

¹⁸ Suikkanen (2020), p.26. Suikkanen draws the distinction between “constitutive consequences” and “causal consequences”.

is the quantity of wellbeing that would be produced by our action, Utilitarianism is indifferent to the way in which this wellbeing is distributed. Consider, for instance,

Two Children: You have 2 children, one of which is disabled, and you must decide whether you should move to the city or to the suburbs.¹⁹ It would be better for your disabled child to move to the city, and it would be better for your other child to move to the suburb. The quantity of wellbeing they would have in both cases is represented in the following table:

	Disabled child	Non-disabled child
Move to the suburb	16	30
Move to the city	20	25

Here, since the overall quantity of wellbeing would be higher if you moved to the suburb ($16+30=46$) than if you moved to city ($20+25=45$), Utilitarianism tells you to move to the suburb. However, if you move to the suburb, the difference in wellbeing between your two children will be much greater ($30-16=14$) than if you moved to the city ($25-20=5$). Therefore, Utilitarianism is often criticised for being insensitive to considerations of equality.

Act consequentialists have a way to accommodate the intuition that a state of affairs in which your children have 20 and 25 units of wellbeing is better than a state of affairs in which one has 30 and the other one 16, even though the sum of wellbeing is less in the former outcome than in the latter. They can adopt an axiology which attributes intrinsic value to equality.

¹⁹ This example is from Nagel (1979) and is discussed by Parfit (1997).

1-2. Ignorance and false beliefs

Once an axiology is adopted, act consequentialists must decide how outcomes will be evaluated in cases of ignorance or false beliefs. This is especially relevant for real-life decisions, which are likely to be made by agents with imperfect information. Consider the following case:

Drowning Baby: you have the opportunity to save a drowning baby at little cost to yourself, which you do. However, the baby is Adolf Hitler. This means that, unbeknownst to you, you have saved someone who will turn out to be responsible for the death of millions.²⁰

Arguably, any plausible moral theory would deem saving a drowning baby permissible, and even required. This is the case despite the fact that, in most cases, one cannot foresee the indirect consequences of one's act of saving this baby. Yet, in *Drowning Baby*, a version of AC which evaluates acts on the basis of their consequences would deem your act of saving the baby impermissible, even though at the time you performed it you had no way of foreseeing the terrible implications it would have. On the contrary, it seems more plausible to say that, although your act of saving the baby has terrible consequences, it was nevertheless permissible, and even required, given your knowledge at the time you performed it.

Now, consider another case:

Rich Uncle: you want to poison your rich uncle in order to inherit his wealth. To do so, you put in his tea what you believe to be cyanide. However, for some reason, the

²⁰ This example is from Joe Horton.

white powder you put in his tea is in fact sugar, which means that he enjoys his tea unscathed.²¹

Any plausible moral theory would deem this act wrong. Yet, in *Rich Uncle*, a version of AC which evaluates acts by looking at their consequences would deem your act permissible, since you merely gave your uncle a sweet tea. In this case, it seems more plausible to say that, although your act did not have negative consequences, it was nevertheless impermissible.

So, how should we evaluate acts performed with imperfect information? In both these cases, it seems like one should evaluate acts on the basis of the outcome the agent *believes* her act would lead to. Saving a drowning baby should be permissible, even though the agent cannot know how much value or disvalue its life will bring about. On the other hand, attempting to harm someone should be impermissible, even though a lucky course of events occurs such that no one is harmed, independently of the agent's doings. Since morality is supposed to offer practical guidance, it would be somewhat odd to determine the permissibility or impermissibility of actions on the basis of outcomes that are not within the agent's control and knowledge.

Therefore, according to the most plausible version of AC, one ought to do what one *believes* would make things go best. We should thus revise the basic definition of AC we began with in the following way:

²¹ This case and the next two ones are inspired by Parfit's discussion of ignorance and false beliefs in Parfit (2011), pp.150-163.

AC: an act is permissible if and only if the agent believes it would make things go best (that is, if and only if the agent believes it would bring about the best state of affairs).

This new version of AC gets things right in the two cases above. In *Drowning Baby*, since you did not know that the baby you were about to save would later be responsible for a genocide, you believed that your act of saving it would make things go best – it is thus permissible. On the other hand, in *Rich Uncle*, you did not know that the powder you put in your uncle's tea was not cyanide, which means you believed that your act would not make things go best – it is thus impermissible.

1-3. Justified and unjustified beliefs

Now, what about cases in which the belief which grounds the action is not only false, but also unjustified? As an illustration, consider a modified version of *Rich Uncle*:

Voodoo: you believe in magic, and instead of trying to poison your uncle with cyanide, you make a voodoo doll of him and stab it with needles in an attempt to kill him from afar. Of course, your attempt is unsuccessful, and your uncle keeps living his life normally.

In this case, just as in the original one, negative consequences are avoided because your belief that your action will lead to your uncle's death is mistaken. The difference is that, in the original *Rich Uncle* case, your false belief is justified (let us say you had no way of knowing that white powder was in fact sugar), whereas, in *Voodoo*, your belief in the power of voodoo

is unjustified (let us say that it is not supported by evidence). Is this difference morally relevant?

Some act consequentialists might say that it is not, and that making a voodoo doll is as impermissible as attempting to poison your uncle. After all, in *Voodoo*, just as in *Rich Uncle*, you had nothing to do with the fact that your action was grounded in a belief that turned out to be false. In both cases, you intended to kill your uncle, and you did what you thought would.

Other act consequentialists might find the difference between justified and unjustified beliefs morally relevant. They might consider that the permissibility or impermissibility of an act depends not on what the agent actually believes the outcome of her act will be, but on what she *should* believe it will be, given available evidence. Therefore, whereas attempting to poison your uncle is impermissible, making a voodoo doll of him is not, because, given available evidence, you should believe that making a voodoo doll of your uncle will not kill him.

Nothing in the structure of AC favours one way of dealing with unjustified beliefs over the other. However, evaluating acts on the basis of the agents' beliefs about their consequences, no matter how unfounded these beliefs are, leads us astray from the most common forms of AC. I also think that doing so is not desirable. Although, for a plausible moral theory to offer practical guidance, it should take into account the fact that we often have to make decisions with imperfect information, the permissibility or impermissibility of actions cannot be independent of evidence altogether.²² I will thus consider that, according to the most plausible version of AC, the permissibility of an act depends on the outcome the agent should believe her act would lead to, given available evidence, rather than the one the agent actually believes her act would lead to. We thus obtain the following definition:

²² See Harman (2015) on the idea that ignorance does not exculpate.

AC: an act is permissible if and only if the agent should believe, given available evidence, that it would make things go best (that is, if and only if the agent should believe, given available evidence, that it would bring about the best state of affairs).

1-4. Uncertainty

There is yet another way in which AC needs refining. In many cases, available evidence does not allow one to anticipate with certainty the outcome of one's actions. Consider this example:

The Drug: as a doctor, you have to choose between three different drugs to cure one of your patients – drug A, B or C. You know that either A or B will cure her completely and that the other one will kill her, but you do not know which one. You also know with certainty that drug C will cure her well enough but without bringing her back to full health. Which drug should you choose?²³

In this case, you have no way of knowing which option is the one that will make things go best – that is, which drug will cure your patient completely. Therefore, AC must have a way to deal with risk and uncertainty.

A plausible way of doing so is to determine the expected value of the different options in a choice situation, and to select the one with the highest expected value. To determine the expected value of an act, one first multiplies the value of each of its possible outcomes by the probability of the outcome obtaining, and then sums up these factors. For instance, in our

²³ This case is inspired from Jackson (1991).

example, let us assume that being restored to full health is worth 10, that dying is worth 0, and that being cured well enough without being restored to full health is worth 9. If you give your patient drug A, there is a fifty percent chance that you will kill her and a fifty percent chance that you will cure her completely, so the expected value of this act is $(0.5 \times 0) + (0.5 \times 10) = 5$. The same applies to drug B. However, if you give your patient drug C, you are certain (probability of 1) that the outcome worth 9 will obtain, which means that the expected value of this act is $1 \times 9 = 9$. Therefore, option C has the highest expected value, and you ought to give your patient drug C.

Note that the act with the highest expected value is not the one which will make things go best. In our example, the act which would make things go best would be giving your patient the right drug, instead of C. However, given the limited information available, simply saying that you ought to choose the option which will lead to the best state of affairs is unhelpful. On the contrary, saying that you ought to choose the option which has the highest expected value preserves the consequentialist idea that what is of ultimate value are states of affairs, while providing practical guidance to agents with imperfect information.

1-5. Final formulation

Given all these modifications, we now have the following view:

AC: an act is right if and only if, given available evidence, the agent should believe it would make things go expectably best (that is, if and only if, given available evidence, the agent should believe it would maximise expected value).

This is the version of AC I will have in mind for the rest of this dissertation (even if, for reasons of space, I mention the shortest definition).

There are four main reasons to accept AC, besides its simplicity. The first one is that it is intuitively compelling, to the point of sounding almost trivial. It seems hard to argue that one should not make things go best or that, if anything is of value, it should not be maximised.

The second reason to accept AC lies in an analogy with rationality. On the one hand, rationality seems to require that one should make things go best for oneself. On the other hand, morality can be seen as a form of extended, impartial rationality. If we accept these two independently plausible assumptions, we must accept that morality requires that one makes things go best from an impartial point of view – impersonally.

The third reason to accept AC is that, in many cases, the implications of AC track widely shared moral intuitions. Imagine, for instance, that you face a choice between saving from death one or two individuals. AC can easily account for the common intuition that you ought to save two people instead of one: other things being equal, a state of affairs in which two people live is better than a state of affairs in which one person lives. Likewise, when you have the possibility to help someone cross the road safely at little cost to yourself, a plausible version of AC can easily make the common-sense prescription that you ought to help them cross the road. More generally, AC's plausibility comes from the fact that, very often, the consequences of our actions do seem to matter. As Joel Kupperman puts it, AC "advocates acting responsibly: instead of adopting the attitude 'I made a moral choice; it fell to earth I know not where', one adopts the attitude of caring about the results of one's action".²⁴ This seems to be an advantage over deontological theories, for which an action can be permissible irrespective of its consequences.

²⁴ Kupperman (1981), p.306.

A fourth reason in favour of accepting AC is that AC can accommodate a great deal of disagreement. As we have seen earlier, it is possible for act consequentialists to adopt different axiologies. This means that AC can be endorsed by people who have very different ethical views, so that these views can be seen as different versions of AC rather than opposing theories.²⁵ Moreover, this allows act consequentialists to address some of the objections raised against their theories, and which I will present in the next parts.

2. Objection 1: Williams on Integrity

Since AC attributes value to states of affairs, it does not matter who brings it about. There is no distinction between positive and negative responsibility – that is, between what I do and what I allow or fail to prevent. This, for Bernard Williams, makes AC incompatible with what he calls “integrity”.²⁶ Consider two of his examples:

Chemist: you are an unemployed chemist with young children to take care of, and your lack of employment has damaging results on them. Someone offers you a well-paid job in a laboratory doing research into chemical and biological warfare. When you say that you cannot take this offer because you are opposed to chemical and biological warfare, you are informed that if you do not take the job, someone else will, and this someone happens to be very enthusiastic about biological and chemical warfare and will do the job zealously. Should you take the job?

²⁵ This idea that most first-order ethical views have a consequentialist equivalent is sometimes called the “consequentialising project”. See Dreier (2011).

²⁶ Williams (1973), pp.93-100.

Guest: as you are visiting a town, you come across a situation in which twenty innocent people are about to get executed by sanguinary soldiers. The captain tells you that he is willing to grant you a guest's privilege: if you kill one of these people yourself, the nineteen others will be set free. Should you kill one person?

According to AC, in *Chemist*, you should take the job: it is very likely that things would be worse in general if the over-zealous person took the job rather than you, and your children would be worse off. In *Guest*, you should kill someone, since the state of affairs in which one person dies is better than a state of affairs in which twenty people die. These implications, Williams argues, threaten the agent's integrity.

In what follows, I distinguish and consider several arguments grouped under the heading of the integrity objection. These arguments concern four different things which AC is said to overlook: the distinction between doing and allowing, one's moral self-conception, self-concern and personal projects, and personal relationships. I argue that, whereas the first two do not constitute significant challenges to AC, the other two do.

2-1. Doing and allowing

For Williams, these two examples suggest that something is wrong with AC. He argues that many would disagree with AC in *The Chemist* and would not find AC's answer as obvious as consequentialists take it to be in *The Guest*. These intuitions are accounted for by the value we attribute, or should attribute, to the distinction between doing and allowing. What matters are not only states of affairs, but also *who* brings about these states of affairs. AC is mistaken because it neglects the idea that "each of us is specially responsible for what *he*

does, rather than what other people do”.²⁷ To make things go best, the act consequentialist agent might be required to perform actions which would threaten her integrity instead of letting things happen, which would preserve her integrity. Therefore, the difference between producing chemical weapons or letting someone else produce them, or between killing someone or letting someone get killed, is morally significant.

However, the importance of the distinction between doing and allowing is not straightforward. Consider this variant of our *Rich Uncle* example:

Cousin: you are visiting your rich uncle, and as you get to the kitchen you see your cousin putting what you know to be poison in his tea. You hide behind the door and stay put. Your cousin gives your uncle the poisoned tea, and your uncle dies.²⁸

Here, you did not kill your uncle; you merely allowed his death to happen. Yet, it is unclear to what extent this distinction is morally relevant, or to what extent your “integrity” is preserved. After all, your uncle is dead, and you could have prevented it. It seems like saying that you did not kill him yourself would not be enough to get you off the moral hook. Admitting that killing your uncle yourself would have been worse, you still hold some degree of responsibility for failing to prevent his death.

Now, let us think again about Williams’ original examples, starting with *Guest*. Although letting the sanguinary soldiers kill twenty people is not as bad as killing twenty people yourself, it still seems pretty bad. Perhaps it is even bad enough so that you ought to kill one person instead – this option does not seem all that crazy. In *Chemist*, pretty much the same

²⁷ Ibid., p.99.

²⁸ This example is inspired by Rachels (1975).

can be said. The overzealous person who would take the job if you do not will probably provoke the deaths of many more innocent people, so that letting them do this when you could prevent it might be at least as bad as provoking the death of some of these people yourself. Given that your children will also be better off if you take the job, this option does not evidently appear as the wrong one. Therefore, it seems like the integrity objection needs more than the distinction between doing and allowing to get off the ground.

2-2. Moral self-conception

In our two examples, it is possible that you will feel bad for performing the action AC endorses – taking a job in a firm conducting research in chemical and biological warfare in *Chemist* and killing an innocent person in *Guest*. From a consequentialist perspective, these negative feelings are only one of the consequences which you should take into consideration in your deliberation. They do not count for much, and they are unlikely to outweigh the positive consequences of your action. But, for Williams, these feelings matter beyond their consequentialist value because “our moral relation to the world is partly given by such feelings, and by a sense of what we can or cannot ‘live with’”.²⁹ By asking us to ignore such moral feelings and by dismissing them as “squeamishness”, he argues, AC alienates one from oneself and undermines one’s integrity. The distinction between doing and allowing matters, because there is something morally problematic in acting against one’s moral self-conception.

Again, we need to say more for this objection to have a grip. Even if we greatly value integrity (understood as respect for one’s moral self-conception), this does not mean that it should override other considerations. When one has a choice between sacrificing one’s arm

²⁹ Williams (1973), pp.103-104.

to save the lives of twenty persons or do nothing, it seems like one ought to sacrifice one's arm, even though physical integrity is of value. Why would it be any different for moral integrity? It seems implausible that integrity, understood as the respect of one's moral self-conception, should always outweigh other considerations. Even if we think it is valuable, in both *Chemist* and *Guest*, there seems to be a number of victims such that we would regard it impermissible to preserve one's integrity instead of saving lives.

This seems pretty straightforward in *Guest* – even Williams does not suggest that you should not kill one person yourself to save the nineteen others. However, things are muddier in *The Chemist*: because of the way the case is designed, the difference between your two options might seem marginal. If you do not take the job, someone else will; and your scope for sabotaging the firm's endeavour is probably very limited. This means that your feelings and the impact taking this job would have on your well-being and integrity might well outweigh the good consequences of taking the job, which supports Williams' claim. However, if, on the contrary, we amend the example to increase the difference between the number of people who would die if you took the job and the number of people who would die if the over-zealous scientist took the job, his claim seems less plausible. Hence, even in this example, there seems to be a number of victims such that we would regard it as impermissible to preserve one's moral self-conception instead of saving lives.

This idea that moral integrity matters at least to some extent can easily be accommodated by AC: act consequentialists could attribute integrity a high value in their axiology, albeit not an absolute one. This way, AC would account for the intuition that one's moral self-conception matters, without making the implausible claim that it matters above everything else.

2-3. Over-demandingness

Even if act consequentialists amend their axiology to account for the value of moral integrity, the fact that it is just one thing of value among others still means that the agent will often have to act against her moral self-conception. But this act consequentialist injunction to disregard the agent's commitments for the sake of an impartially good state of affairs goes even further.

We have seen that, since AC attributes value to states of affairs, it disregards the distinction between negative and positive responsibility – one is responsible for much more than what one does. This means that, when we evaluate options to determine which one is the most valuable and ought to be chosen, we have to take into account all the available alternatives, even those which are not directly under consideration. Importantly, this is the case for every single action we perform.

In Williams' words, AC places virtually no limits to "one's apparently boundless obligations" to bring about the best state of affairs.³⁰ All aspects of our lives are subject to the moral obligation to make things go best for everyone. This is implausibly demanding, because most of our daily actions are suboptimal in comparison to some alternatives. For instance, having coffee with a friend is impermissible because I could be volunteering at a soup kitchen instead, which would probably make things go better. Likewise, taking tennis lessons is impermissible because I could be giving this money to charity. Therefore, AC is incompatible with most personal projects and commitments – moral or not – which have to be sacrificed for the sake of an impartially good state of affairs.

The projects and commitments which the agent is morally expected to forego to maximise good consequences include self-concern. Just as each and every one of the agent's projects

³⁰ Ibid., pp.109-110.

and commitments are evaluated for their costs and benefits for everyone, the agent's interests are to be weighed impartially against everyone else's. For instance, to determine the permissibility of getting a job, I ought to take into account the negative consequences this might have for the person I was in competition with. If it turns out that these outweigh the positive consequences it has for me, AC tells us it is impermissible for me to accept the job.

In fact, according to AC, one should dedicate one's life to making things go best, which would require one to give up on any comfort or goods beyond that necessary to one's survival and ability to maximise value. Considering the current state of the world – in which many have close to nothing – act-consequentialist morality would require that I give away everything I own, since this would make things go best by vastly improving the situation of many. It might even prescribe that I lay down my life, if my continuing existence turned out to affect them negatively to a large enough extent – which it probably does. Needless to say, these extremely demanding prescriptions are highly counterintuitive.

One consequentialist response is to say that people do a better job at looking after their interests than others, so things go better if everyone looks after themselves more than they look after others. However, this is far from being enough. All areas of one's life are now part of the trade-off with others' interests. Even taking into account the fact that people know best what is good for them, which means that consequences will be better if they give some priority to satisfying their interests, AC will still end up being extremely demanding.

We are starting to grasp the significance of William's objection. Even if, in cases such as *Guest* and *Cousin*, one's obligations plausibly extend beyond what one does, there seems to be many cases in which this should not be the case. The problem is that AC denies this. Since it demands of the agent that she places every aspect of her life – her moral self-conception, her personal projects, her wellbeing – on a par with the interest of others, it is

incompatible with self-concern and personal commitments. It leaves no areas of our lives untouched and deems impermissible most of our actions, even the most harmless ones. As such, it is implausibly demanding.

2-4. Special relationships

Act consequentialists could push back and say that, for all we know, it could be the case that morality is very demanding, and that AC's over-demandingness is not a decisive argument against it. Elizabeth Ashford thus argues that the source of AC's demandingness is not internal to the theory; rather, it is the fact that, in the world as it is, "there are continually persons whose vital interests are threatened and, given modern communications, the relatively well-off are continually able to help them".³¹ Given the current state of the world, any plausible moral theory is likely to entail deep conflicts between its dictates and the agent's personal commitments. However, in a conceptually and practically realisable state of the world, in which few would find themselves in such situations of emergency, AC would not conflict with personal projects and commitments anymore.

Nevertheless, one might still worry about the overly demanding implications of AC in the real world. Moreover, the problem with AC's incompatibility with personal projects goes beyond over-demandingness. Williams argues that AC neglects in a more fundamental way certain commitments that give meaning to our lives, such as personal relationships. This is because, even if we grant Ashford's response, it remains the case that, in a world in which far lesser people are in a situation of emergency, AC's dictates would still coincide with what special relationships demand only *coincidentally*. Of course, one worry is that, in the world as it is, AC does demand that we betray our partner or sacrifice our children for the sake of

³¹ Ashford (2000), p.430.

an impartially best state of affairs. But the problem is more insidious: even in a state of the world in which AC did not entail very demanding obligations, it would only be out of luck that the act consequentialist agent would not have to go against the demands of her special relationships.

Consider the following example:

Rescue: in a shipwreck, a man has a choice between saving a stranger or his wife. After deliberating briefly, he concludes that it is permissible for him to give priority to his wife. So, he saves her.

Here, it might well be that a plausible version of AC, which attributes value to personal relationships, would deem the man's action permissible. The problem is that, even then, AC would still exclude the possibility of valuable personal relationships such as the one between this man and his wife.

The thought is that the structure and reasoning of AC is incompatible with the value of special relationships – or, rather, with the behaviours and attitudes which the respect for such value demands. The fact that AC requires the agent to go through a consequentialist calculation to know whether saving his wife is permissible (whether it is at the time of the action or not) is incompatible with the existence of special relationships, since these require a form of unconditional partiality. For Williams, such situations seem to “lie beyond justification”. When the man rescues his wife because he concluded it was permissible for him to give her priority, he has already had “one thought too many”.³²

³² Williams (1981), p.18.

We have now distinguished a second facet of Williams' integrity objection. Although it is related to the demandingness objection, it remains distinct. The problem with AC's disregard for the identity of the agent is not only that it leads to overly demanding obligations. It is also that it prevents, in all possible states of the world, the existence of special relationships. Yet, such relationships are what gives meaning to our lives. In Williams' words, "unless such things exist, there will not be enough substance or conviction in a man's life to compel his allegiance to life itself", or to give meaning to morality.³³ Since it seems plausible that special relationships are a morally important aspect of our lives, this is a significant objection to AC.³⁴

2-5. Summary

I have outlined what I take to be the four main aspects of Williams' Integrity Objection. The first two – that AC fails to make a distinction between doing and allowing, and that AC neglects the importance of one's moral self-conception – do not seem very persuasive. However, the last two – that AC is over-demanding, and that AC is incompatible with special relationships – constitute significant challenges to AC.

³³ Ibid., p.17.

³⁴ Railton (1984) addresses this objection by theorising what he calls "sophisticated consequentialism", which is the idea that, to make things go best, it is highly plausible that the agent must act for other motives than to make things go best. Constantly thinking about impartially good consequences is self-defeating, since certain things (such as special relationships) are impartially valuable and yet require partial commitments. Hence, it might well be that, in order to make things go best overall, one must cultivate special relationships, even if those often require that one does not engage in consequentialist calculations, and sometimes that one acts in suboptimal ways.

This response is compelling, but it takes us away from AC. We are not evaluating acts anymore; we are evaluating the consequences of motives, inclinations or dispositions. This is closer to Rule Consequentialism, which is another theory I discuss in chapter 2.

3. Objection 2: Rights and Duties

Another one of the main objections to AC I want to present concerns rights and duties.³⁵ We have seen that AC is accused of being too demanding. It allows only for a minimal degree of self-concern and demands that the agent sacrifices personal projects and commitments for the sake of bringing about an impartially best state of affairs. In the same way, this commitment to impartially best states of affairs also entails that the agent must sacrifice the projects and commitments of others, including their lives. Just as her interests are to be weighed against everyone else's, so are the interests of the individuals affected by her actions. This means that acts which common-sense morality condemns are not only permissible, but even required, as long as they maximise overall value.

Consider this example:

Surgeon 1: a surgeon has five patients who will die if they do not get an organ transplant soon. A sixth patient is booked in for a benign surgery. During her operation, the surgeon kills this patient and takes from her the organs necessary to save the five other patients, who live as a result.

In this case, some versions of AC suggest that your act was permissible, because a state of affairs in which five live is better than a state of affairs in which one lives. This result is, quite obviously, rather counterintuitive. Many would argue that the patient who gets killed has a right to live, and to receive medical care without being murdered.

Act consequentialists have a way to mitigate this objection. As we have seen in part 1, their axiology can accommodate the idea that certain acts can have value in themselves –

³⁵ See McCloskey (1965).

or, in the case of killing, disvalue. In this case, they can argue that the disvalue of killing one person is greater than that of five persons dying of natural causes, so that the act of the surgeon is impermissible.

However, this response can only take us so far. Even if we attribute disvalue to the act of killing, there will be a number of patients the surgeon could save by killing one person so that, according to AC, the benefits of her action would outweigh the costs. On the contrary, what objectors argue is that the patient has a right not to get murdered *no matter the consequences*. The concept of rights seems to require that they cannot be trumped by other considerations *at all*.

Act consequentialists can bite the bullet and maintain that there is a number of victims such that it would be permissible to murder someone to spare them. This number can be high enough for this claim to be plausible, and for AC to get the intuitive result that the surgeon should not murder her patient in *Surgeon 1*.

But, even so, AC is not out of trouble, as the proposed solution faces problems of its own. Consider a modified version of the previous example:

Surgeon 2: A surgeon is in the process of saving five patients from cardiac arrest. Their cardiac arrests were prompted by natural causes and, if she attends to all of them at the same time, she can save their lives. A new patient, who also suffers from cardiac arrest, is brought into another room. Unlike the five other patients, the new patient's cardiac arrest was caused by someone poisoning her such that, if she dies, she will have been murdered. Since she is in a different room from the other five, the surgeon has a choice between saving her alone, or saving the five.

In this example, the surgeon has a choice between saving one person from being murdered or saving five people from a natural death. If act consequentialists are committed to saying that one person being murdered is worse than five people dying naturally, which they have to be in order to address the previous objection, they must accept that she ought to save the one person being murdered in *Surgeon 2*. However, it seems plausible that, when facing such a choice, one ought to save the five instead, even if they would die from a natural death.

Therefore, act consequentialists are facing a choice. They must either accept that, in *Surgeon 1*, the surgeon ought to kill one patient in order to save the five, or that, in *Surgeon 2*, she ought to save the one from murder instead of saving the five from a natural death. In both cases, it seems like they cannot avoid implausible conclusions. So, AC has trouble accommodating the concept of moral rights.

Unsurprisingly, the same thing can be said about duty. We have already seen that AC's dictates contradict the demands (and thus the duties) involved in special relationships. But AC's disregard for duties extends further. Just as with rights, duties matter only to the extent that they make things go best. For instance, one ought to respect one's promises, or to tell the truth, only as long as doing so would make things go best.

Here, the same type of reply as the one offered in *Surgeon 1* can be made. Act consequentialists can attribute disvalue to the act of lying, in which case lying would be permissible only when the overall benefits are large enough – say, when a life is at stake. This seems like a plausible implication. However, once again, AC face difficulties in cases in which one can either prevent one lie from being told by someone else or bring about good consequences short of saving a life – say, if they could save someone from paraplegia. If they want AC to have the plausible implication that it is impermissible to lie unless a life is at stake, they must accept that they ought to prevent a lie from being told rather than save

someone from paraplegia, which seems implausible. Therefore, act consequentialists cannot avoid their theory having implausible conclusions, and AC has difficulties accommodating the concept of duties.

4. Objection 3: Collective Impact

There is yet another category of acts which common morality regards as demanded of us, and which AC is unable to account for. This problem is known as the problem of collective impact. Consider the following case:

Coffee: You see a documentary about the horrors of the coffee industry, including the exploitation of farmers and the use of slave and child labour. The next time you order a coffee, you ask for a Fairtrade brand. However, you find that the Fairtrade coffee is not as good as the one you usually buy, which is not Fairtrade. So, the next time you get a coffee, you order your usual brand.³⁶

Many would think that your act of buying a coffee which you know has been produced under miserable working conditions is impermissible. The problem is that AC is unable to accommodate this judgement. Although things would be best if everyone stopped buying non-Fairtrade coffee, your purchasing this cup of non-Fairtrade coffee does not make a difference. After all, as Julia Nefsky puts it, “you are a mere speck in a sea of coffee drinkers”.³⁷ The coffee company will not stop exploiting farmers or change anything about

³⁶ This example is from Nefsky (2015).

³⁷ Ibid., p.267.

their mode of production simply because you bought a cup of Fairtrade coffee instead. Therefore, since drinking non-Fairtrade coffee brings you more satisfaction, and since it does not make a difference to global exploitation, AC would prescribe buying it.

AC has similarly counterintuitive implications in other cases involving collective impact. Consider, for instance, changes in one's consumption aimed at reducing one's carbon footprint in an attempt to contribute to countering global warming. Since such acts require one to give up on some comfort, and yet have an impact close to null when taken in isolation, they will not be endorsed by AC. Instead, AC prescribes free riding on the efforts of others, even if fighting global warming requires everyone to perform just this kind of acts, because the cumulative effect would be tremendously good. Therefore, AC cannot account for the permissibility or obligatoriness of acts which do not make a difference when taken in isolation, but which are nevertheless generally thought to be demanded of us.³⁸

5. Objection 4: Aggregation

The act consequentialist idea that what is of moral value are impartially good states of affairs is intuitively compelling, plausible and simple. However, we have seen that this entails a neglect of rights and duties, and that our interests, commitments, and those of the ones we love are to be impartially weighted against everyone else's, leading to implausible conclusions. Now, we will see that this idea that individuals' interests are all part of a vast

³⁸ Shelly Kagan (2011) argues that AC can address the problem of collective harm if we consider that there is a threshold at which repeated actions of one sort will have catastrophic consequences. Since we do not know when this threshold will be met, we should look at the expected (dis)utility of performing one such act, which is likely to be negative. However, Nefsky (2012) has argued persuasively that appealing to such "triggering cases" does not solve the problem.

trade-off aiming at maximising impartially good consequences gives rise to another objection – one regarding the issue of aggregation.

Since AC aggregates people's interests to look at overall consequences, it fails to make a distinction between one person's suffering immensely and a large number of people being mildly indisposed.³⁹ As an illustration, consider the following example:

Death vs Broken nails: You have a choice between saving one person from death or sparing a large number of people a broken nail. Is there a number N such that you ought to spare N people a broken nail rather than save one person from death?

According to AC, the answer is yes. If the number of people you could spare a broken nail is large enough, you ought to help them rather than save someone from death, because doing so would make things go impartially best. This is rather counterintuitive.

Note that no change to the axiology could address this objection. Act consequentialists might mitigate it by attributing extremely high value to saving a life, but this would not fully solve the problem. There will always be a number of victims risking a broken nail such that one ought to spare them rather than save someone else's life. On the other hand, to prevent this conclusion, act consequentialists could assign no disvalue to a broken nail. However, they would then have to accept the implausible implication that we have no moral reason to spare people a broken nail, even if we can do so at no cost. In both cases, AC cannot avoid implausible conclusions.

³⁹ This is often called the separateness of persons objection. See Brink (1993), Rawls (1999) pp.23-27 and Nagel (1970) pp. 138-142.

However, it is not clear how significant this objection from aggregation is, considering the problems faced by available alternatives. On the one hand, non-aggregative views, which prescribe that we always prevent the gravest harm, are at least as counterintuitive as AC.⁴⁰ Consider the following example:

Death vs Paraplegia: You have a choice between saving one person from death or saving a large number of people from paraplegia. Is there a number N such that you ought to save N people from paraplegia rather than save one person from death?

According to non-aggregative views, the answer is no. There is no number of people you could save from paraplegia such that you ought to save them rather than save someone from death. This is at least as implausible as AC's implications in *Death vs Broken nails*.

An intuitive alternative is to adopt a partially aggregative view, according to which there is no number N such that one ought to spare N people a broken nail rather than save a life, but there is a number N such that one ought to save N people from paraplegia rather than save a life.⁴¹ Such a view would prescribe saving one individual from death in *Death vs Broken nails*, no matter how large the number of people you could spare a broken nail is, but saving a large enough number of people from paraplegia rather than saving one person from death in *Death vs Paraplegia*.

This seems like the most plausible option. However, although they are intuitively appealing, partially aggregative views face many problems of their own. For instance, it has been argued that these views cannot avoid counterintuitive implications in cases involving

⁴⁰ For a defence of non-aggregative views, see Anscombe (1967/2005) and Taurek (1977).

⁴¹ For defences of partially aggregative views, see Kamm (1993), Scanlon (1998) and Voorhoeve (2014) and (2017).

risk, mixed groups or two-stage decisions.⁴² Therefore, it is not clear whether they fare better than AC regarding the problem of aggregation.

Finally, the two other moral views I will consider in chapters 2 and 3 – Rule Consequentialism and Rule Contractualism – also seem to be unable to avoid aggregative conclusions.⁴³ Therefore, although the problem of aggregation seems to constitute a significant challenge against AC, it is not a decisive one. Other competing views, such as Rule Consequentialism and Rule Contractualism, suffer from it as well; and the views that avoid this challenge face equally important challenges as a result. For this reason, I will leave the issue of aggregation aside in my evaluation of Rule Consequentialism and Rule Contractualism in the next chapters.

Conclusion

After looking at the problems posed by false beliefs, ignorance and uncertainty, I have argued that the most plausible version of AC is the following one:

AC: an act is right if and only if, given available evidence, the agent should believe it would make things go expectably best (that is, if and only if, given available evidence, the agent should believe it would maximise expected value).

⁴² See Tomlin (2017), Horton (2017), (2018) and (2020) for a thorough overview of these problems.

⁴³ See Otsuka (2000) who argues that Scanlonian Contractualism has to take into account the claims of groups of individuals in order to reach the partially aggregative conclusions Scanlon defends.

This moral view is supported by strong arguments: it is simple and intuitively appealing, it corresponds to the independently plausible idea that morality is about making things go impersonally best, and it can accommodate many first-order ethical intuitions.

AC also faces a number of serious problems. Of all the objections we have looked at, the objection from aggregation is probably the less worrying. After all, it is not clear how other views can avoid aggregative conclusions, or whether those that do fare better, considering the problems they face as a result. However, AC is not out of trouble. Williams' integrity objection, once unpacked, constitutes a strong challenge, as it exposes this view as implausibly demanding and incompatible with special relationships. We have also seen that AC neglects rights and duties, and that it has counterintuitive implications in collective action cases.

All these objections give us good reason to search for a better moral view. In the next chapter, I ask whether Rule Consequentialism can do better.

Chapter 2

Rule Consequentialism

My aim in this dissertation is to carve out a conceptual space for a hitherto neglected moral theory, Act Contractualism. I do this by looking at the merits and limitations of three much more familiar moral theories: Act Consequentialism, Rule Consequentialism and Rule Contractualism.

We have seen in chapter 1 that Act Consequentialism (hereafter AC) is vulnerable to a number of important objections. In this chapter, I look at another moral view, Rule Consequentialism (hereafter RC) and ask whether it fares better. I look at RC for two reasons. First, it has been formulated as an answer to some of AC's shortcomings. As such, it is an influential and promising alternative to AC. Second, introducing RC will make it easier to introduce Rule Contractualism, which I discuss in the next chapter. This will in turn make it easier to introduce and evaluate Act Contractualism.

According to

Rule Consequentialism (RC): an act is permissible if and only if it conforms to the best set of rules, where the best set of rules is the set such that things would go better (or at least as well) if everyone complied with this set, or accepted this set, than if everyone complied with any alternative set, or accepted any alternative set.

The significance of the distinction between compliance versions of this view and acceptance versions of this view will be addressed later.

I argue for two main claims in this chapter: that RC seems able to address at least some of the primary objections raised against AC, and that it is vulnerable to another set of interrelated objections.

I start by recapitulating in section 1 what I take to be the main objections to AC, and showing in section 2 how RC avoids some of them. The following six sections are dedicated to presenting objections to RC and evaluating Brad Hooker's defence of this moral view, which is the most comprehensive attempt to answer these objections.⁴⁴

In section 3, I present the Collapse Objection, according to which RC collapses into AC and is therefore vulnerable to all the same problems. I then present Hooker's response to this objection, which consists in adopting a Universal Acceptance version of the theory, rather than a Universal Compliance version. We will see that, if RC avoids the Collapse Objection, it seems to become vulnerable to another criticism, the Inconsistency Objection. However, this is mistaken. As Hooker argues, RC is not committed to maximising good consequences; it is therefore not inconsistent when it fails to do so. This answer to the Inconsistency Objection will serve to introduce Hooker's reflective equilibrium argument in favour of RC. This defence is grounded in the theory's supposed alignment with widespread moral intuitions and rests on the claim that a unifying non-consequentialist rationale can be offered for them. I will examine both parts of this argument in turn.

I will start with the alleged similarity between RC's prescriptions and common intuitions. I will show in section 4 that RC has counter-intuitive implications in the face of a challenge known as the Ideal World Objection (hereafter IWO). I will then consider amendments to RC which address the IWO by narrowing the gap between our considered moral judgements and RC's prescriptions.

⁴⁴ Hooker (2002).

However, RC is not yet out of trouble. In section 5, I will restate the challenge in light of the previous responses to the IWO. The objection now concerns the second part of Hooker's reflective equilibrium argument: rule consequentialists must provide a unifying rationale for the widespread intuitions which their theory appears to track.

In section 6, I will evaluate Hooker's claim that the rationale behind RC is that we ought to act in ways that are impartially defensible. I will argue that no substantive account of impartial defensibility can justify following RC. Therefore, I conclude that Hooker's reflective equilibrium argument in favour of RC fails, which gives us good reason to retreat into AC or to look for a better view.

1- Objections to AC

Remember that, according to

Act Consequentialism (AC): an act is an act is permissible if and only if, given available evidence, the agent should believe it would make things go expectably best (that is, if and only if, given available evidence, the agent should believe it would maximise expected value).

AC is simple, intuitively compelling and plausible. However, it faces a number of important problems. I will now recapitulate what I take to be the main objections to AC, before asking whether RC is able to answer them.

1-1. Over-demandingness

AC is accused of being overly demanding.⁴⁵ It requires the agent to impartially weigh her wellbeing and personal commitments against everyone else's. Moreover, according to AC, all aspects of our lives are placed under moral scrutiny. No act – however insignificant – is left untouched. This means that most of our daily actions, which we usually take to be at least permissible, are deemed impermissible by AC.

As an illustration, consider the following example:

Pleasant afternoon: You have an afternoon off from work. You have always wanted to learn how to play tennis, so you decide to get a private tennis lesson. After that, you go for a coffee with a friend.

Here, it seems pretty uncontroversial that, by spending your afternoon in this way, you did nothing wrong. As long as you did not promise anyone not to play tennis, as your friend is not a morally dubious individual, or as no other such conditions which would constitute a moral reason against engaging in these activities apply, it should be permissible to do so. However, AC is likely to deem playing tennis and having coffee with a friend impermissible. This is because these acts are suboptimal compared to available alternatives: things would go best if you gave the money you spent on the tennis lesson to a charity, or if you volunteered at a local foodbank instead of spending time with your friend.

AC is thus incompatible with any degree of self-concern beyond that necessary to stay alive and keep bringing about good consequences. In the current state of the world, it requires that we give away everything we own, and that we spend as much time as we can, to improve

⁴⁵ See Williams (1973), pp.109-110.

the situation of others. It might even require that we lay down our lives if it turned out to lead to an impartially best state of affairs. This makes AC implausibly demanding.

1-2. Special relationships

AC is also accused of being incompatible with special relationships.⁴⁶ By requiring that not only the agent's interests and commitments, but also those of her loved ones, be weighed against everyone else's, AC is incompatible with the demands associated with such relationships. Consider the following example:

Rescue: in a shipwreck, a man has a choice between saving a stranger or saving his wife. After deliberating briefly, he concludes that it is permissible for him to give priority to his wife. So, he saves her.

Even if a plausible version of AC would deem this man's action permissible, it would not avoid this objection. As Bernard Williams puts it, the problem is that the act consequentialist agent has "one thought too many".⁴⁷ In this case, the fact that the man goes through consequentialist calculation before saving his wife, or that he might be required to justify his action afterwards, prevents the kind of unconditional partiality that is needed for special relationships to exist. Since special relationships are valuable, this is a significant problem for AC.

⁴⁶ See Williams (1981).

⁴⁷ Ibid.

1-3. Rights and duties

Just as AC demands that the agent sacrifice her personal projects and commitments for the sake of bringing about the impartially best state of affairs, AC requires that she sacrifice the personal projects and commitments of others, and even their lives. Consider this example:

Surgeon: a surgeon has five patients who will die if they do not get an organ transplant soon. A sixth patient is booked in for a benign surgery. During her operation, the surgeon takes organs from this patient and takes from her the organs necessary to save the five other patients, who live as a result.⁴⁸

Even if a plausible version of AC attributes high disvalue to the act of killing, there will always be a number of patients such that the surgeon ought to kill another patient to save them. This goes against common-sense morality and the intuitive idea that the one patient has the right to live and to receive medical care without being murdered. More generally, AC has trouble accommodating the concept of moral rights, since it can only account for them whenever respecting them makes things go best. The same goes for duties: according to AC, one ought to tell the truth, or respect one's promises, only as long as doing so makes things go best.

1-4. Collective impact

Another category of acts which common-sense morality regards as demanded of us, and which AC cannot account for, are those whose effects are felt only when they are performed

⁴⁸ This example is from Foot (1978) and is discussed at length by Thomson (1985).

by a large number of people. This is known as the problem of collective impact. Consider the following case:

Coffee: You see a documentary about the horrors of the coffee industry, including the exploitation of farmers and the use of slave and child labour. The next time you order a coffee, you ask for a Fairtrade brand. However, you find that the Fairtrade coffee is not as good as the one you usually buy, which is not Fairtrade. So, the next time you get a coffee, you order your usual brand.⁴⁹

Here, AC would deem your act permissible. Since the effect of you buying a single cup of coffee is negligible, and since you prefer the non-Fairtrade coffee, things go best if you buy the non-Fairtrade coffee.

This is counter intuitive. Most of us believe that buying coffee which you know has been produced under miserable working conditions is impermissible. Even if your act alone does not change anything, it is because of consumers like you that coffee farmers keep being treated so poorly. If many consumers refrained from buying non-Fairtrade coffee, the firm producing it would go under or would be forced to improve the conditions of its farmers. Yet, AC is unable to account for this idea because it only looks at the consequences of one act of buying non-Fairtrade coffee. AC has similarly implausible conclusions in other collective impact cases, such as those real-life cases in which we are together bringing about environmental catastrophe.

⁴⁹ This example is from Nefsky (2015).

2- Does RC address these objections?

We have four main objections to AC: its over demandingness, its incompatibility with special relationships, its neglect of rights and duties, and the problem of collective impact.

Does RC address these objections?

2-1. A consequentialist rationale behind the first three objections

Joel Kupperman suggests that the first three objections to AC we have just discussed do not disprove consequentialism—that is, they do not disprove the simple idea that consequences are what ultimately matter. Rather, these objections are themselves consequentialist, because they still come from a general concern for good consequences.⁵⁰ For instance, if we criticise AC for not prescribing promise keeping or truth telling, and if we believe that promise keeping and truth telling are valuable, it is because they have good consequences. Although this suggestion seems odd at first sight, on reflection it has some plausibility. As Kupperman puts it, “if the world were such that promise keeping and other forms of fidelity to commitments brought about misery and other bad consequences, would we (or should we) still prize such a character?”⁵¹

This argument can be made about all the things of value which AC is said to neglect – permissible self-concern, special relationships, and rights and duties. Consider, for instance, the objection that AC is too demanding. There seem to be good consequentialist reasons to allow for important degrees of self-concern. A world in which everyone sacrifices everything they have to maximise impartial value seems to be self-defeating, and certainly less valuable

⁵⁰ Kupperman (1981).

⁵¹ Ibid, p.308.

than a world in which everyone looks after himself in such a way as to be able to live a good life and sustain consequentialist practices over time.

Consider next the objection that AC is incompatible with special relationships. If we think it is wrong to be cold and calculating when considering our obligations to people we are involved in special relationships with, it can be argued it is because doing so generally has negative consequences. A world in which these types of relationships exist is better than a world in which they do not, and this can easily be expressed in consequentialist terms.⁵² Therefore, a plausible form of consequentialism should accommodate the fact that it is better for people to show partiality towards those they are involved in special relationships with.

Lastly, consider the objection that AC neglects rights and duties. Again, there seem to be good consequentialist reasons to consider rights and duties as “trump cards” which should be respected irrespective of the consequences. Doing so might be best overall, as it would protect us from special-pleading and slippery slopes.⁵³

Therefore, the first three objections – that from over-demandingness, special relationships, and rights and duties – do not necessarily undermine the appeal of consequentialist evaluations. Rather, they point towards the necessity of promoting rules, policies, motives or tendencies that make things go best, rather than promoting individual acts that make things go best. This suggests that a plausible moral theory – one which would deal well with the first three objections – might well be consequentialist. Such a theory, instead of evaluating acts directly by looking at their consequences, would evaluate them indirectly by looking at their conformity with rules, policies, motives or tendencies selected for their consequences.

⁵² See Railton (1984).

⁵³ Kupperman (1981), pp.309-310.

2-2. Over-demandingness

Remember that, according to

RC: an act is permissible if and only if it conforms to the best set of rules, where the best set of rules is the set such that things would go better (or at least as well) if everyone complied with this set, or accepted this set, than if everyone complied with any alternative set, or accepted any alternative set.

RC thus evaluates acts indirectly: their permissibility or impermissibility is determined by the consequences of the rules that permit them. In light of Kupperman's reading of the objections from over-demandingness, special relationships, and rights and duties, we begin to see how RC might fare better than AC with respect to these challenges.

Regarding the objection from over-demandingness, it seems plausible that things would go better if everyone followed a set of rules which makes room for reasonable degrees of self-concern rather than a set of rules which does not allow self-concern. This means that, according to RC, acts which show some reasonable degree of self-concern are permissible. Moreover, giving small amounts of time and money would be enough to alleviate global poverty if *everyone* contributed, which means that RC plausibly requires us to contribute *some* of our time and money without being overly demanding.

Therefore, since things would go best if everyone showed a reasonable degree of self-concern and contributed to some extent to alleviate global poverty, RC seems to get the right result in cases such as *Pleasant Afternoon*. Paying for a tennis lesson and having coffee with a friend is permissible, because things would go best if everyone followed a set of rules which

allowed such harmless activities to be performed, and which prescribed occasionally altruistic behaviour such that no one would be required to give more than reasonable amounts of their time and resources. RC thus avoids the over-demandingness objection.

2-3. Special relationships

Let us turn to the second objection – that AC is incompatible with special relationships. Saying that things would go best if everyone followed a set of rules according to which one ought to give priority to one's loved ones will not do. Even if it means that RC would prescribe the intuitively plausible course of action in cases such as *Rescue*, this does not get to the heart of the problem.

Remember that the objection against AC was not simply that it could not account for the permissibility of showing some degree of partiality toward one's loved ones – we have seen in chapter 1 that some tweaking of the act consequentialist axiology could solve this problem. Rather, the significance of the objection rests in the fact that the very structure of AC, and the idea that everything is to be weighed against other costs and benefits, is incompatible with the existence of special relationships, which (according to Williams) should lie beyond justification.

It is plausible that the best set of rules would be one that tells us to blindly devote ourselves to the ones we love without thinking about the permissibility or consequences of such attitude. If such devotion is necessary for special relationships to exist, and if special relationships are part of the best state of affairs, the best set of rules is likely to permit this kind of blind devotion. In *Rescue*, the man would thus be permitted, and even required, to save his wife without asking himself whether this would be permissible.

Yet, there is a sense in which the rule consequentialist agent might still be said to have “one thought too many”. The very existence of a rule regarding the permissibility of the blind devotion involved in special relationships seems to go against the idea that these relationships lie beyond justification. Moreover, the fact that this rule is based on a consequentialist calculus – the fact that the permissibility of blind devotion lies in a cost-benefit analysis in which its value is weighed against that of everything else – is incompatible with the incommensurable value Williams seems to attribute special relationships.

Still, I believe RC addresses well enough the objection from special relationships. Not only does it make it more likely that showing partiality for one’s loved ones will be permissible, but it also allows the kind of unconditional partiality Williams suggests is necessary for special relationships to exist. Admittedly, the consequences of such unconditional partiality are still being weighed against other costs and benefits, which suggests that they do not lie beyond justification. However, since blind devotion is permitted (and even required), the rule consequentialist agent does not problematically engage in consequentialist calculus. Although the consequentialist justification of special relationships which occurs at a higher level might fail to acknowledge their incommensurable value, it does not seem incompatible with their existence.

2-4. Rights and duties

The third objection – that from rights and duties – is easily addressed by RC. Rights and duties can be seen as rules which would make things go best if everyone followed or accepted them. For instance, it is plausible that things would go best if everyone followed a set of rules which would forbid killing, prescribe telling the truth and respecting promises, and so on. Hence, it seems like RC gets the right result in cases such as *Surgeon*.

2-5. Collective impact

Now, to the last objection. According to AC, what matters are the consequences of a single act. This explains why this view cannot account for the obligatoriness of acts which have close to no impact when taken in isolation, but which would have a great impact if they were performed by many people. On the other hand, RC asks us to consider the consequences rules would have if *everyone* complied with them, or accepted them. In other words, it takes into account the cumulated consequences of the universal performance of certain acts. It is thus well equipped to deal with the problem of collective impact.

For instance, it is plausible that things would go better if everyone followed a set of rules which would prescribe recycling, or which would forbid buying products made under poor working conditions, than if everyone followed a set of rules which did not. Therefore, unlike AC, RC would get the intuitive result in *Coffee*. It would prescribe buying the Fairtrade coffee, because this act is permitted by the set of rules which would make things go best if it was accepted or followed by everyone.

Importantly, the rationale behind this conclusion also seems on point. According to RC, it is impermissible to buy the non-Fairtrade coffee because things would go worse if everyone followed a rule allowing such acts. This seems to get to the bottom of what we feel is going on in collective action cases, which is often expressed by the question “what if everyone did like you?”.⁵⁴

⁵⁴ Hooker (2002), pp.4-5.

3- The Collapse Objection and the Reflective Equilibrium Argument

3-1. The objection

We have seen that RC is immune to at least some of the main objections raised against AC. It accounts for the permissibility of self-concern and special relationships, for the existence of rights and duties, and deals appropriately with the problem of collective impact. RC thus seems to track common intuitions more accurately and fares better than AC with respect to these objections. However, RC is vulnerable to another set of interrelated objections, which will be the object of the rest of this paper.

The first of these objections, the Collapse Objection, runs as follows.⁵⁵ If a rule R does not prescribe the act which has the best consequences in a certain situation S, it should be modified so that it does, lest it cannot be the rule which has the best consequences when everybody acts upon it – that is, when it is universally followed. This will also be the case for situation S*, S**, and so on. Since we need R to be the rule which has the best consequences in each and every situation, it must contain exception clause after exception clause. We end up with a rule which prescribes acting in such a way as to maximise good consequences in each instance, which is what AC prescribes.

As an illustration, consider the rule “do not break a promise”, along with the following example:

Letter. You have promised your sister you would post a letter for her before the post office closes. On your way to the post office, you see a child drowning in a lake. No

⁵⁵ See Smart (1973) pp.10-12, Gert (2005) pp.214-218 and Podgorski (2018) p.281.

one else is around. You could save the child but, if you did, you would not make it to the post office before it closes.

In this case, following the rule “do not break a promise” would arguably make things go worse than not following this rule; a state of affairs in which a child drowns is worse than one in which you do not fulfil your promise to post a letter. However, there is another rule which would prescribe the act which would make things go best in this case: “do not break a promise unless by doing so you would save the life of a child”. If everyone followed this rule, which contains an exception clause regarding cases such as *Letter*, things would go better than if everyone followed the original rule without an exception clause, since some people might find themselves in a situation similar to our example, in which an exception would make things go best.

Now, consider the following example:

Birds: You have promised your neighbour to chop down a tree in her garden. However, you now realise that the last two specimens of an almost extinct bird species are nesting in this tree. You cannot reach them to put them out of harm’s way. Chopping down the tree would kill them.

In this case, following the rule “do not break a promise unless by doing so you would save the life of a child” would not make things go best; a state of affairs in which a species goes extinct seems worse than one in which you fail to fulfil a promise to chop down a tree. Once again, there is another rule which would prescribe the act which would make things go best in this situation: “do not break a promise unless by doing so you would save the life of a child

or avoid killing the last specimens of an endangered species”. This rule would have better consequences when followed by everyone than the previous one, because some people might find themselves in a situation similar to that in *Birds*, in which an exception would make things go best.

We now get the gist of the argument: for every situation in which making an exception to the rule “do not break a promise” would make things go best, the rule should be amended so that it prescribes the act which makes things go best. Only then can this rule be the one which makes things go best when it is followed by everyone, in every situation. Therefore, the best rule, according to RC – that is, the rule which makes things go best when it is followed by everyone – is the one which tells us to make things go best in each and every situation. In other words, it is a rule requiring us to follow AC.

Since RC is extensionally equivalent to AC, the argument goes, its commitment to an indirect evaluation of acts via rules seems superfluous. It results in a more complex theory which has the same implausible implications as AC, and which does not really address the objections to AC we have considered in the previous section.

Consider, for instance, the objection from right and duties. In order for a rule forbidding killing to be the one which has the best consequences when it is followed by everyone in every situation, it should make an exception every time deviating from the rule would make things go better than sticking to it. This is the case in our *Surgeon* example: things would go better if five people lived rather than one – that is, if the surgeon did not obey a rule forbidding killing. Therefore, the rule which has the best consequences when it is followed by everyone, in every situation (including in situations such as *Surgeon*), is a rule which tells us to kill one patient to save five. This means that the surgeon ought to follow this rule, and thus to kill her patient.

Since this implication is exactly the one which grounds objection from rights and duties against AC, RC does not address the objection from right and duties. The same goes for all the other objections to AC we have looked at in the previous section: since RC is extensionally equivalent to AC, it runs into all the same problems.

3-2. Response: Compliance and Acceptance

So far, I have talked about rules which would make things go best if they were “accepted or followed” by everyone, as if the two were interchangeable. However, they are not, and the rule consequentialist response to the Collapse Objection hangs on this distinction. The Collapse Objection only holds if we assess rules on the basis of the consequences they would have if they were universally *followed* by everyone – that is, if we consider the consequences of universal *compliance* with rules. It can be avoided by versions of RC which assess rules on the basis of the consequences they would have if they were universally *accepted* by everyone – that is, if we consider the consequences of the universal *acceptance* of rules, as Brad Hooker suggests.⁵⁶

Hooker argues that looking at the consequences of universal compliance with rules is a mistake. The universal adoption of a moral code does not simply result in its being universally followed; it comes with externalities, such as that associated with the publicity, implementation and internalisation of this moral code. Therefore, the most plausible version of RC – Universal Acceptance RC – should select rules on the basis of the consequences they would have if they were universally accepted, rather than if they were universally followed. Importantly, since the publicity, implementation and internalisation of a rule have costs and benefits beyond that of the universal performance of the acts prescribed by that

⁵⁶ Hooker (2002), pp.75-80.

rule, the rules which have the best consequences according to Universal Acceptance RC are not extensionally equivalent with AC.

Consider for instance the rule “do not break promises except when doing so would make things go best”. Since it consists in maximising good consequences in each individual instance, it is extensionally equivalent to AC. Moreover, since this rule would have the best consequences when *followed* by everybody, Universal Compliance RC prescribes following it. Therefore, we can see that, at least with respect to this rule, AC and Universal Compliance RC are extensionally equivalent. However, Universal Acceptance RC would not prescribe acting on this rule, because the consequences of its being *accepted* by all are different from its being followed by all, with the former unlikely to be good. For instance, if everyone accepted a rule requiring us to break promises whenever doing so would make things go best, that would eventually result in a situation where people cannot rely confidently on others. Since this state of affairs would be worse than one in which people could still trust each other, Universal Acceptance RC would not endorse this rule.

In this example, the bad consequences of the collapse of trust in society result not from individuals *complying* with this rule, but from public awareness of the rules’ exception clauses – that is, from individuals *accepting* this rule.⁵⁷ This suggests that a Universal Acceptance version of RC leads to different results than a Universal Compliance version. The former does not prescribe acting on the rule which would make things go best if it was universally followed. Therefore, it is not extensionally equivalent to AC, and it avoids the Collapse Objection.

The Universal Acceptance version of RC vindicates our earlier responses to the objections to AC. Consider again *Surgeon*: if everyone *accepted* a rule according to which it is permissible for doctors to kill some of their patients to save other patients, no one would

⁵⁷ See Hooker (2005) p.269-270.

seek medical help anymore, which would have disastrous consequences. Since public awareness of this rule would result in a state of affairs which is worse than a state of affairs in which people can confidently seek medical help without fearing for their lives, Universal Acceptance RC would not endorse it. Therefore, Universal Acceptance RC, by avoiding the Collapse Objection, once again addresses the objection from rights and duties.

Similar points can be made about the other objections: by looking at the consequences of the universal *acceptance* of rules regarding self-concern, special relationships and collective impact, rule consequentialists can avoid the problematic implications which prompted objections to AC.

3-3. The Reflective Equilibrium Argument

It has been argued that, if rule consequentialists resist the Collapse Objection, their theory becomes inconsistent.⁵⁸ If, like Hooker, we insist on the extensional non-equivalence between RC and AC, we must accept that acts prescribed by RC will sometimes have suboptimal consequences, which is rather odd for a theory which intrinsically values good consequences. As Gary Watson puts it, “if what justifies the rule or the right are just the effects on well-being of general compliance, then the effects on well-being should be decisive in deciding whether or not to comply in particular cases”.⁵⁹ No matter the rule consequentialist’s preferred axiology, her theory seems to fail by its own standards when it prescribes acting in ways which do not bring about the best state of affairs.

Hooker has answered this objection, I believe successfully.⁶⁰ RC does *not* imply an overarching commitment to maximising the good, but only to acting in ways consistent with

⁵⁸ Smart (1956); Card (2007); Willand (2010); Rajczi (2016).

⁵⁹ Watson (1998), p.172.

⁶⁰ Hooker (2002) pp.100-102 and (2007).

rules which would maximise the good if they were universally accepted. Although rules are selected according to whether their internalisation would maximise the good, acts are not evaluated in this way. They are only evaluated “by reference to the rules thus selected”.⁶¹ Therefore, RC is not inconsistent when it prescribes acts which do not maximise the good.

Hooker makes clear that RC is not grounded in consequentialist considerations. Instead, he argues, “the best argument for rule-consequentialism is that it does a better job than its rivals of matching and tying together our moral convictions, as well as offering us help with our moral disagreements and uncertainties”.⁶² This reflective equilibrium argument consists of two parts. First, RC accords with many of our firmly held moral convictions – unlike AC and deontological accounts which depart from them in important ways. Second, RC is grounded in a rationale which justifies its prescriptions, without which it could not help us with our moral disagreements and uncertainties.⁶³ In other words, the best argument for RC, Hooker tells us, is that it is both plausible and explanatory. In the remainder of this paper, I will evaluate both parts of this reflective equilibrium argument in favour of RC, starting with the first – the supposed plausibility of RC’s conclusions.

4- The Ideal World Objection

4-1. The objection

RC asks us to follow rules which would make things go best if everyone accepted them. However, in a world in which there is no such universal acceptance, following these rules

⁶¹ Hooker (2002), p. 101.

⁶² Ibid.

⁶³ I will say more about this requirement in section 5.

might lead to disaster. Consider a rule forbidding violence.⁶⁴ Since things would go best if *everyone* accepted this rule, RC prescribes following it. Yet, in a world like ours, where some individuals do use violence to oppress others, pacifism would have disastrous consequences. Although most people intuitively believe that using violence to prevent someone from killing innocents is at least permissible (if not required), RC seems unable to account for this permission.

At first sight, it seems like this challenge could be addressed by adopting conditional rules. Our pacifist rule could become “do not use violence, unless other people do, in which case use proportionate violence to contain them”. The problem is that RC has no means of selecting this new rule over any other conditional rule starting with the same antecedent. For instance, the rule “never use violence unless other people do, in which case go crazy and kill everyone” fares just as well under RC, because no one would use violence if *everyone* accepted it. Therefore, RC is unable to discriminate between two such conditionals, or between them and the original non-conditional rule, as long as it asks us to consider the consequences of their universal acceptance, which are the same. By evaluating rules in ideal circumstances of full compliance, RC thus fails to prescribe the right rules in non-ideal conditions of partial compliance. This is the Ideal World Objection (IWO), as formulated by Derek Parfit and Abelard Podgorski.⁶⁵

4-2. Response: General Acceptance

To address the IWO, rule consequentialists have amended RC to prescribe rules in situations of partial adherence.⁶⁶ Hooker thus suggests that we consider the consequences

⁶⁴ This example is from Parfit (2011) pp. 312-320 and is discussed by Podgorski (2018) pp.281-281.

⁶⁵ Parfit (2011) pp.312-320 and Podgorski (2018).

⁶⁶ These attempts are outlined in Podgorski (2018).

a rule would have if it was accepted by 90% of the population, rather than by the whole population.⁶⁷ He argues that doing so provides us with rules addressing non-compliance while still upholding the idea that the right moral code is the one which would be best suited “for acceptance *by the whole society of human beings*”.

However, although this strategy deals with non-compliance and thereby reduces the gap between RC and widespread intuitions in cases like our pacifism example, it leaves RC vulnerable to variants of the IWO. Indeed, a set of rules selected for its good consequences at a 90% acceptance level, even if it provides guidance for some degree of non-compliance, might not address situations which would occur at different acceptance levels.⁶⁸ Following rules on the basis of their consequences at any fixed acceptance level may thus lead to disaster.

Consider this case:

Gremlin: X is a set of rules such that, when accepted by 90% of the population, it has better consequences than any alternative set. However, an all-powerful Gremlin will destroy the world if you act upon X while X is accepted by less than 90% of the population.⁶⁹

Since things would go better if 90% of the population accepted X than if they accepted any other set of rules, Hooker’s version of RC would prescribe following X in the real world, no matter the actual acceptance level of X. However, if, in the real world, less than 90% of the

⁶⁷ Hooker (2002), pp.80-85.

⁶⁸ Ridge (2009), pp.68-71.

⁶⁹ This example is inspired by Rosen (2009).

population accept X, this would lead to the destruction of the world by our Gremlin.⁷⁰ Therefore, RC might prescribe that we act in ways which, at real-world acceptance levels, would have disastrous consequences. Note that this applies to any variant of RC that considers fixed acceptance rates.⁷¹

Attempts to immunise RC against the IWO by evaluating the consequences of rules at multiple acceptance rates also fail. Consider Parfit's strategy: selecting rules which would make things go best if they were followed "by any number of people".⁷² As Michael Ridge argues, "the idea that there would be one code which is best for each and every one of these [possible acceptance levels] would seem *prima facie* to be extremely unlikely".⁷³ Whether or not such a code exists is an empirical matter, and Parfit's solution implies that, in a world in which no code fares better than others at all acceptance levels, nothing is morally required. This cannot be right.

Ridge thus proposes to look at a rule's average consequences for all levels of acceptance.⁷⁴ This ensures that one moral code would be selected while avoiding the arbitrariness associated with selecting a fixed rate. However, this solution still does not immunise RC against the IWO. As Podgorski argues, our evil Gremlin could guarantee that marvellous things would happen if we acted upon the set of rules X at any level of acceptance except the actual one, at which he would destroy the world.⁷⁵ Since the consequences of this set of rules would be the best on average, Ridge's RC would prescribe following it in the

⁷⁰ It might be objected that this Gremlin example is too far-fetched and has little bearing on real-world cases. However, it is only a dramatic way of making real-world phenomena more vivid. See Podgorski (2018) on real-world exemplifications of what he calls "utility mines". However, for the sake of the argument, I will concede this point in section 5.

⁷¹ This includes Smith's (2010) Optimum Acceptance Level RC, which offers to select rules on the basis of their consequences at their optimal level of acceptance.

⁷² Parfit (2011) pp.312-320.

⁷³ Ridge (2009) p.67.

⁷⁴ *Ibid.*, p.72.

⁷⁵ Podgorski (2018), p.287. See also Hooker and Fletcher (2008), p.349.

actual world, even if this meant provoking a disaster. Therefore, no matter what level of acceptance rule consequentialists adopt for the evaluation of rules, following the rules prescribed by RC can lead to disaster in the actual world.

4-3. Response: Avoid Disaster

RC threatens to depart in important ways from the common intuition that one should avoid provoking disasters whenever possible. We have seen that it has extremely counter-intuitive implications in fanciful cases such as *Gremlin*, but it is also the case in more realistic cases. Consider for instance

Boat: You are on a boat with nine other passengers. Nearby, another boat capsized, and there are nine people in the water about to drown. There are nine lifejackets on your boat, which you and the other passengers could throw overboard to save the nine people from drowning. However, the lifejackets are stored in individual compartments with capricious locks, and opening them is slightly painful and annoying. Therefore, none of the other passengers of your boat bother doing anything. What should you do?⁷⁶

Here, it seems obvious that you should open all the compartments and throw nine lifejackets to the people drowning, even if no one else is doing anything, and even if opening all the locks might end up being quite painful. However, a version of RC which prescribes following the rules which have the best consequences when accepted by 90 percent of the population

⁷⁶ This case is inspired by Hooker (2002), p.164.

would tell you to throw only one lifejacket overboard.⁷⁷ This is because things would go best if 90 percent of the passengers on your boat followed the rule “throw one lifejacket overboard”: nine passengers would only have to open one lock and experience minimal discomfort, and nine lives would be saved. Therefore, 90 percent acceptance RC would tell you to save only one life. This is so even if, in the real world, less than 90 percent accept this rule and no one else will throw lifejackets overboard.⁷⁸

Since it asks us to conform to rules which have the best consequences at idealised acceptance rates, RC might prescribe that we act in ways which lead to disasters in cases such as *Gremlin* and *Boat*. However, Hooker insists that RC does not have such counter-intuitive implications, because the best set of rules contains an overriding Avoid Disaster rule.⁷⁹ In *Boat*, for instance, you should keep throwing lifejackets overboard, because the rule “avoid disaster” overrides the rule “throw one lifejacket overboard”.

Two questions emerge. First, how should one avoid disaster? There might be many ways to do so, and RC should specify which one is permissible. The most plausible interpretation of this provision is to consider that one should act in ways which are as close as possible to the spirit of the rule that one is violating to avoid disaster. For instance, to avoid disaster when others do not comply with pacifism, one should use as little violence as possible, but not wreak havoc. In *Boat*, avoiding disaster presumably means saving the nine from drowning.

⁷⁷ Hooker (2002, p.106) argues that the rule regarding principles of aid which would make things go best is in fact “do good for others up to some threshold of aggregate self-sacrifice” (see also Hooker and Fletcher 2008, pp.347-348). However, since this rule has the same consequences than the rule I suggest when it is accepted by 90% of the population, it is unclear why Hooker’s suggested rule would be the one selected by RC.

⁷⁸ We have seen in part 2 that RC avoids the over-demandingness objection by demanding that individuals give only the amount of time or money which would be enough to make things go best if everyone contributed. Counter-intuitive implications such as the ones in *Boat* are the other side of the same coin: by looking at what would happen if (almost) everyone contributed to provide aid, RC might not be demanding enough.

⁷⁹ Hooker (2002) pp.98-99. See also Brandt (1988) pp.150-151 and (1989), pp.87-88.

This leads us to the second question: what counts as a disaster? The lower the bar, the more we deviate from rules whenever they have bad consequences, and the closer we get to extensional equivalence with AC.⁸⁰ As we have seen, rule consequentialists resist this collapse. Hooker argues that the rules which would have the best consequences if they were accepted by most of the population cannot have too many exception clauses nor be too ambiguous: the internalisation of complex rules would be too costly, and the implementation of ambiguous rules would leave the door open to special pleading and interpretation.⁸¹ This is why he insists that avoiding disasters does not mean following AC. Despite the “avoid disaster” clause, RC still prescribes acting in ways which do not have the best consequences in each situation.

To see this, consider the examples of the previous section and replace the disastrous consequences of acts prescribed by the different versions of RC by “bad-although-not-disastrous consequences”. Or imagine that, in *Boat*, the nine people in the water would not drown but be injured – just not enough to count as a disaster. Whether we adopt fixed-rate RC or multiple-rates RC, the rules which are selected can easily be the ones which have bad-although-not-disastrous consequences in the actual world. No alteration to the level of ideality we are considering will change that.

Furthermore, RC can require that we act in ways which have close to no impact in the real world. For instance, imagine our Gremlin would bring about the best consequences possible if 90% of the population performed their best impression of a chicken at 12pm. RC would then require us to act like a chicken at 12pm, even though in the real world it would be entirely pointless. Although this is far from a disaster, it is a rather odd implication.

⁸⁰ See Arneson (2005) ft6 and Podgorski (2018) p.288.

⁸¹ Hooker (2002) pp.94-99.

Therefore, despite the Avoid Disaster rule, RC is still vulnerable to a variant of the IWO: in all its versions, our moral evaluation of acts does not rest upon what happens in the actual world but on what would happen if the rules permitting these acts were accepted by a certain proportion of the population. What happens in the real world has a very limited influence on the selection of the rules to be followed. This leads to counter-intuitive results: individuals may be required to act in ways which would lead to consequences short of disaster, or in ways which do not make the tiniest difference in the real world.

5- The Challenge Restated

This variant of the IWO need not be the death knell of RC. Since it rests on the idea that RC is not sensitive enough to consequences in the real world, it is only fatal on consequentialist grounds. Rule consequentialists could bite the IWO bullet and claim that morality is simply not concerned with good consequences in the real world. This is improbable, however, as most of them concede that some regard for real-world consequences is justified. The fact that they even address the IWO by amending the theory to prevent disasters suggests that they do care about consequences in the actual world.⁸² Moreover, these concessions are vital since, without them, RC would lose its claim to tracking widely held moral intuitions. Therefore, rule consequentialists are more likely to say that, although they do care about bringing about good consequences in the real world, this is not an overriding aim.⁸³ Doing so would be a way of rejecting the IWO by arguing that, although RC fails to maximise good consequences, it does so for the sake of another overriding

⁸² Card (2007) p.249.

⁸³ That is the essence of Hooker's (2007) reply to Card.

commitment. This idea brings us to the second part of the reflective equilibrium argument – that RC is supported by an explanatory rationale.

Nothing we have said so far addresses this part of the argument, nor strikes a fatal blow to the first part – that RC does a better job than other theories at tracking widely-held moral intuitions. The amendments to RC we have considered in section 4 narrow the gaps between RC and widespread moral intuitions that the original version of the IWO had identified. Moreover, one could argue that cases such as our Gremlin example are too far-fetched and that they do not disprove RC. In real-world (Gremlin-free) scenarios, the difference between the consequences of following a rule in a world with idealised rates of acceptance and the consequences of following them in the actual world is likely to be more moderate. If we place the bar for what constitutes a disaster relatively low, rule consequentialists can still maintain that their theory does not depart from widely held intuitions to unreasonable extents.

Although I believe that real-world phenomena resemble the Gremlin case – pacifism and aid are two of them –, and although I find the implications of RC in these cases extremely implausible, I will concede this point for the sake of argument. Let us grant that RC still tracks common moral intuitions reasonably well. After all, we have seen that RC accommodates a reasonable degree of self-concern and has plausible implications in collective action cases. Moreover, it accounts for widely held intuitions regarding deontological considerations – it forbids lying, stealing and killing. With Hooker’s amendments, RC also accounts to some extent for our general concern for good consequences. Therefore, it arguably does a better job at tracking common moral intuitions than AC or deontological accounts.

However, granting the first half of the reflective equilibrium argument does not get RC off the hook. It is not enough for a theory to have intuitively plausible implications; we also want it to be explanatory. As Hooker puts it, moral theories must not only systematise pre-

theoretical judgements, they should also “identify a fundamental principle that both (a) provides a unified explanation of why our more specific considered moral convictions are correct and (b) justifies them from an impartial point of view”.⁸⁴ Recall the reflective equilibrium argument I have presented in section 3. If it is to offer guidance in cases of disagreement and uncertainty, an appropriate theory must be able to tell us which of our pre-theoretical moral intuitions are justified, and which ones we ought to revise. To fulfil these twin justificatory and revisionary functions, it must offer a unifying explanation for its prescriptions – one that justifies them.

Moreover, let us not forget that RC does have some counter-intuitive implications. We have seen that it sometimes prescribes that we act in ways which would bring about bad-although-not-disastrous consequences. This is not a problem in itself: we usually expect moral theories to prompt us to revise some of our intuitions. However, an appropriate moral theory should be able to justify these revisions. In the case of RC, it seems like we are owed a fairly strong justification for why we ought to revise our intuitive judgements regarding cases such as the ones addressed in section 4. For instance, we want to know *why* we should revise our intuition that we should avoid acting in ways which bring about very-bad-although-not-disastrous consequences, or that we are not morally required to act like a chicken when absolutely no good will come out of it.

The need for a unifying rationale is even more pressing now that rule consequentialists such as Hooker have addressed the IWO in ways which suggest that RC is not merely concerned with adherence to the best set of rules, but that it is *also* concerned with consequences in the real world. Indeed, the Avoid Disaster rule cannot be justified on other grounds than that following it would make things go better in the real world than following the

⁸⁴ Hooker (2007), p.516.

best set of rules (that is, the set of rules which has the best consequences at an idealised acceptance level).⁸⁵ This is because, when a set of rules is selected because it has the best consequences at a certain acceptance level (say, 90 percent), it already provides the best rules for all the kinds of situations which would occur at this acceptance level. If situations which qualify as disasters were happening at this acceptance level (say, if one person would not be rescued in a case like *Boat*), then the best set of rules should be amended so that it addresses these situations, lest it would not be the best set of rules in ideal circumstances. Therefore, at an ideal acceptance level, the set of rules that tells you to avoid disasters does not have better consequences than the set of rules that does not tell you to avoid disasters, because this code is such that disasters *do not occur* at this ideal acceptance level. This is why the Avoid Disaster rule can only be justified on the grounds that it makes things better in the real world.⁸⁶

As Hooker told us in his answer to the Inconsistency Objection, this does not imply that rule consequentialists have accepted an overarching commitment to maximising the good. However, this idea that considerations pertaining to consequences in the real world are *sometimes* decisive calls for an explanation consistent with the rest of the theory. Without it, the rule consequentialist's answer to the IWO (and, more precisely, the Avoid Disaster rule) threatens to expose RC as a hybrid theory crafted ad hoc around common intuitions, and without any explanatory power.

⁸⁵ Contra Hooker (2002) p.98.

⁸⁶ Rule consequentialists might reply that a code which contains clauses for every potential disaster would be too complex, which would make its internalisation too costly, and that therefore a code containing an Avoid Disaster rule would have better consequences. However, if we place the bar for what constitutes a disaster relatively high, there might not be that many clauses to internalise. It also seems like a code which contains an Avoid Disaster rule would leave the door open to special pleading and interpretation, and that its vagueness would therefore constitute a significant cost. Therefore, it is not obvious that an Avoid Disaster rule would be part of the ideal code.

Even if you disagree with this reasoning, it is enough for my point if you agree that rule contractualists show *some* concern for real-world consequences. Hooker himself concedes this in his reply to Card: "it would be wrong to suggest that rule-consequentialism pays no attention to the actual consequences of people's following rules" (Hooker 2007, p.518).

The challenge can now be restated: in their answer to the IWO, rule consequentialists must explain, via a unifying rationale, why we should address situations of non-compliance and prevent disaster but *not* maximise good consequences generally.

6- A Rule-consequentialist rationale?

6-1. Impartial Defensibility

Hooker asks, “if the consequences of everyone’s feeling morally free to do a given kind of act would be better than the consequences of everyone’s not feeling free to do it, how can acts of this kind be wrong?”⁸⁷ As we have seen, such acts can be wrong on consequentialist grounds, since they will probably fail to bring about the best consequences in real-world circumstances. However, what is suggested by Hooker is that these acts are nevertheless right, albeit for non-consequentialist reasons.

Hooker, among others, offers a contractualist rationale in favour of RC when he suggests that following its dictates may be the best way to “act in ways that are impartially defensible”.⁸⁸ Impartial defensibility would thus be RC’s overriding commitment. It would justify adhering to rules the general acceptance of which would make things go best *and* having some regard for consequences in the form of an Avoid Disaster provision. Since Hooker does not provide a clear account of impartial defensibility, I will assess the plausibility of this claim according to what I think constitutes the best such account.

⁸⁷ Hooker (2002), p.1.

⁸⁸ Ibid., pp.99-104. Hooker denies that his theory is contractualist because he rejects the Scanlonian Contractualist definition of justifiability and considers impartial defensibility more broadly. However, I maintain that a commitment to justifiability is a contractualist commitment, even if it is not a Scanlonian Contractualist one.

6-2. Following RC in the real world is not impartially defensible

RC asks us to follow an ideal set of rules, which is the set of rules which would have the best consequences in an ideal world of quasi-total acceptance. That following this code in these ideal circumstances would be impartially defensible seems plausible. But why would following it in non-ideal circumstances be equally defensible? Hooker's reasoning is that, since morality is "a collective enterprise, a practice to be shared", a moral code "should be internalised and followed by everyone, not just by you or by me or by any mere sub-group of the whole".⁸⁹ It is hard to disagree that, ideally, everyone should internalise the same moral code. But what does this have to do with guidance for action, which is what morality purports to provide?

Hooker tells us this much: it would be morally best if everyone internalised the same moral code, and if this moral code was the one which had the best consequences in these circumstances of total acceptance. From this, we can infer the following guidance for individual action. *If* an individual could make the universal internalisation of this code a reality, it would be wrong not to do so. Moreover, *if* our following a code would lead to its internalisation by the rest of humanity, we would be morally compelled to follow the code which would be best in circumstances of universal internalisation. Therefore, *if* individuals could bring about the universal internalisation of a moral code, and *if* they could do so by following this moral code themselves, individuals would be morally compelled to follow the code which would have the best consequences when internalised by the whole population. Doing so would be impartially defensible.

⁸⁹ Ibid., p.1.

However, individuals are *not* able to bring about the internalisation of an ideal moral code by others simply by acting on it. Their adhering to it does nothing towards its internalisation by the rest of the population.⁹⁰ Although it is plausible to argue that bringing about the collective internalisation of an ideal code would be impartially defensible, this justification is not available for acting on it in the real world, since doing so does not contribute to this collective internalisation.⁹¹

6-3. Another account of Impartial Defensibility?

One could object to my argument by saying that the conception of “impartial defensibility” I have been considering is consequentialist. Indeed, I have assumed that bringing about the internalisation of an ideal moral code would be impartially defensible. I have further argued that, since following RC would not bring about this collective internalisation, RC is not impartially defensible. In other words, I have assumed that bringing about the best consequences is impartially defensible and argued that following RC is not impartially defensible because it does not bring about the best consequences. However, this is not much of a problem if we take seriously Hooker’s claim that RC is not committed to maximising good consequences anyway. Since the rationale behind RC is non-consequentialist, RC does not aim at promoting the best state of affairs in the real world – that is, it does *not* aim at bringing about the general internalisation of this ideal code. This does not mean that following RC is not impartially defensible. So, can another account of impartial defensibility rescue RC?

One possibility could be to understand this argument in Kantian terms, and to say that following an ideal code in the real world is impartially defensible because it is a requirement

⁹⁰ Or, if it does – for instance by “leading by example” – it does so only marginally.

⁹¹ However, the collective institutions which do have the power to bring about collective internalisation – through education, for instance – should keep this ideal code in sight. See Podgorski (2018) p.292.

of rationality. However, if the impartial defensibility Hooker has in mind is indeed Kantian, we are left to wonder why RC should be concerned with real-world consequences at all. Although the Avoid Disaster rule ensures that RC's implications are less counter-intuitive, it results in an uneasy compromise between deontological and consequentialist considerations, which cannot be reconciled by a fully Kantian rationale.

The same applies to contractualist defences of RC, which consist in claiming that it would be rational or reasonable for everyone (under various ideal conditions) to agree to live by the set of rules that makes things go best.⁹² We have seen that the Avoid Disaster rule would not be part of the best set of rules. Therefore, even if we admit that everyone would reasonably agree to live by the best set of rules, this is not enough to show that RC, as it is formulated by Hooker, would equally be the object of hypothetical contractualist agreement. Although the Avoid Disaster rule makes RC more intuitively plausible, it is not readily accounted for by a contractualist rationale.⁹³

As it stands, the appeal to impartial defensibility cannot rescue RC from the accusations of intuitionism that emerged once the theory was amended to avoid the implausible implications identified by the original version of the IWO. Contrary to what Hooker suggests, the onus is now on the rule consequentialist to offer a substantive account of impartial defensibility to remedy this situation.⁹⁴ In the meantime, the reflective equilibrium argument in favour of RC fails.

⁹² See for instance Harsanyi (1955), Brandt (1979), Parfit (2011) and Levy (2013).

⁹³ The Avoid Disaster rule could be defended by an appeal to real-world justifiability. However, I will argue in the next chapter that we must distinguish what we can reasonably agree to in the real world from what we would reasonably agree to in an ideal world. Since the contractualist defence of RC rests on the idea that we would agree to live by the best set of rules in an ideal world, it does not account for the Avoid Disaster rule.

⁹⁴ See Hooker (2002), p.104.

Conclusion

In the previous chapter, we looked at important objections to AC. In this chapter, we have seen that, with respect to some of these objections, RC fares better than AC. Unlike AC, RC accounts for the permissibility of self-concern and special relationships, and for the existence of rights and duties. It also deals appropriately with the problem of collective impact.

However, I have shown that RC is vulnerable to another set of objections. In response to the first one – the Collapse Objection – rule consequentialists have formulated a Universal Acceptance version of RC, which addresses this objection. I have argued that the challenge is now that rule consequentialists must offer reasons to adopt their view beyond that it does not collapse into AC.

As we have seen, Hooker offers a reflective equilibrium argument in support of RC. We should adopt this view, he tells us, in part because it has intuitive conclusions, but also because it ties these conclusions together by justifying them. I have argued that this reflective equilibrium defence fails.

I have restated the IWO to show that the problem is not simply that RC has counter-intuitive implications because it selects rules to be followed in the real world on the basis of what would happen in an idealised world. After all, rule consequentialists have provided us with ways to mitigate these counter-intuitive conclusions; and RC fares better than AC with respect to the objections covered in chapter 1. Moreover, we *expect* moral theories to question – and sometimes challenge – pre-existing moral intuitions. Few consider common-sense morality to be an absolute benchmark, and it is generally accepted that an appropriate moral theory will prompt certain revisions.

The real problem brought out by the IWO is that RC is unable to do that. If it is to fulfil this revisionary function, a moral view must be grounded in an explanatory rationale, and we have

seen that rule consequentialists have not yet provided us with one. Their response to the original version of the IWO puts them in the difficult position of having to provide a unifying rationale that explains why we ought to act according to rules which would make things go best if they were generally accepted and *yet* make exceptions when real-world consequences are negative enough. As it stands, the rationale offered by Hooker – that RC allows one to act in impartially defensible ways – is unsatisfactory and cannot account for the hybrid prescriptions of RC. Therefore, even admitting that RC does a better job than its rivals at tracking widespread moral intuitions, it remains a poor candidate for reflective equilibrium because it cannot justify these intuitions via a unifying rationale. Even if it is plausible, it has no explanatory power.

Without an explanatory rationale, we have no reason to accept RC's counterintuitive implications, or even to feel justified in holding the pre-theoretical judgements which RC does track accurately. We have no reason to adopt RC, except that it avoids the counter-intuitive conclusions of AC. However, this comes at too high a price. In resisting the collapse into AC, rule consequentialists render RC vulnerable to the IWO, which is inescapable because it relates to a fundamental structural feature of RC – its indirect evaluation of acts via rules. Since the IWO means that RC is either extremely counter-intuitive or that it has no explanatory power, we seem to have good reason to retreat back to AC, and to focus our efforts towards addressing the problems this view faces.

There is another view, however, which might do better than RC in addressing the objections to AC while avoiding the IWO. This view – Rule Contractualism – will be the object of the next chapter.

Chapter 3

Rule Contractualism

My aim in this dissertation is to carve out conceptual space for a hitherto neglected moral theory, Act Contractualism. I do this by looking at the merits and limitations of three much more familiar moral theories: Act Consequentialism, Rule Consequentialism and Rule Contractualism.

We have seen in chapter 1 that, although it is appealing, Act Consequentialism (hereafter AC) is vulnerable to a number of important objections. In chapter 2, we have seen that Rule Consequentialism (hereafter RC) fares better than AC in relation to at least some of these objections. However, RC is vulnerable to another challenge, the Ideal World Objection (hereafter IWO), which relates to a fundamental structural feature of RC – its indirect evaluation of acts via rules.

It might seem that we have good reason to retreat back to AC and to focus our efforts on addressing the problems this view faces. However, there might be another way to go. In this chapter, I ask whether another moral theory fares better than RC in addressing the objections raised against AC while avoiding the IWO. This theory is

Rule Contractualism: an act is permissible if and only if it is allowed by a set of rules the universal acceptance of which is appropriately justifiable to everyone.

I start in section 1 by introducing Rule Contractualism in more detail. In section 2, I look at the similarities and differences between Rule Contractualism and RC. I argue that, since

Rule Contractualism resembles RC in that it relies on a process of idealisation/universalisation, it seems well-equipped to address at least some of the objections raised against AC, in much the same way that RC addresses them. The main difference between these views is that Rule Contractualism is concerned with the justifiability of rules, whereas RC is concerned with the value of rules. I argue that this difference constitutes an appealing motivation for adopting Rule Contractualism instead of RC. However, I contend in section 3 that the similarities between RC and Rule Contractualism also mean that Rule Contractualism, too, is vulnerable to a version of the IWO. In section 4, I thus introduce a possible alternative to Rule Contractualism: Act Contractualism. I argue that the IWO gives us good reason to adopt Act Contractualism instead of Rule Contractualism. In the last part, I briefly compare Act Contractualism with AC, and I argue that, although Act Contractualism faces some significant challenges, it is worthy of serious consideration.

1. Rule Contractualism

1-1. Rule Contractualism in the literature

According to

Rule Contractualism: an act is permissible if and only if it is allowed by a set of rules the universal acceptance of which is appropriately justifiable to everyone.

Rule Contractualism is familiar in the moral theory literature. Perhaps the most well-known version of this view is the one defended by Tim Scanlon.⁹⁵ According to

Scanlonian Contractualism: An act is wrong “if its performance would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement”.⁹⁶

Here, acts are assessed indirectly, via principles (or via what we might more standardly call rules).⁹⁷ Their permissibility depends on the justifiability of the principles that permit them – where appropriate justifiability is understood in terms of reasonable unrejectability. Moreover, the reasonable unrejectability of principles is determined on the basis of what would happen if they were generally accepted.⁹⁸ These two aspects of Scanlon’s view – the evaluation of rules in terms of their justifiability and the reliance on a process of universalisation – suggest that Rule Contractualism is a plausible interpretation of Scanlonian Contractualism.

The view defended by Nicholas Southwood is also a version of Rule Contractualism.⁹⁹ According to

Deliberative Contractualism: An act is permissible if it is allowed by a common code which we would agree to live by if we were perfectly deliberatively rational.¹⁰⁰

⁹⁵ Scanlon (1998).

⁹⁶ *Ibid.*, p.153.

⁹⁷ In this chapter, I will use rules and principles interchangeably.

⁹⁸ Scanlon specifies that “when we are considering the acceptability or rejectability of a principle, we must take into account not only the consequences of particular actions, but also the consequences of general performance or nonperformance of such actions and of the other implications (for both agents and others) of having agents be licensed and directed to think in the way that the principle requires”. This means that we are looking at the consequences of the general acceptance of principles, rather than mere general compliance. See Scanlon (1998) pp.202-203

⁹⁹ Southwood (2010) and (2019).

¹⁰⁰ Southwood (2010), p.86. A more complete formulation is found in Southwood (2019), p.531: “It is morally impermissible/obligatory for A to X iff if we were all perfectly deliberatively rational and charged with the task of agreeing upon

Here again, acts are assessed indirectly; their permissibility depends on their conformity with a code (that is, a set of rules). And, again, a set of rules is selected for its justifiability – where appropriate justifiability is determined by whether a given code would be the object of hypothetical agreement. Moreover, since the object of hypothetical agreement is a *common code* by which *everyone* would live – that is, a common code which everyone would comply with, but also accept – we are again required to look at the consequences of the general acceptance of a given set of rules to determine whether it would be justifiable.¹⁰¹ The only difference with Scanlonian Contractualism is that justifiability is not expressed in terms of reasonable rejection but rather in terms of hypothetical agreement between perfectly rational individuals.

Lastly, another familiar rule contractualist view is Derek Parfit's. According to

Kantian Contractualism: “an act is wrong unless such acts are permitted by some principle whose universal acceptance everyone could rationally will”.¹⁰²

Moreover, Parfit argues that principles whose acceptance everyone could rationally will cannot be reasonably rejected, and that such principles are thus justifiable to everyone.¹⁰³ Therefore, once again, the permissibility of acts depends on the justifiability of the rules that permit them – where appropriate justifiability is determined by what everyone could rationally will. Since, once again, justifiability is also assessed on the basis of what would happen if

a common code by which to live, then we would agree to live by a common code, C, that includes a principle P that forbids/requires A to X”.

¹⁰¹ See Southwood (2019), pp.531-532.

¹⁰² Parfit (2011), p.341.

¹⁰³ This is Parfit's Convergence Argument. See Parfit (2011), pp.411-412.

certain rules were universally accepted, Kantian Contractualism is a form of Rule Contractualism as well. (Note that this view relies even more explicitly than Scanlonian Contractualism and Deliberative Contractualism on a process of universalisation.)

To conclude, Scanlonian Contractualism, Southwood's Deliberative Contractualism and Parfit's Kantian Contractualism are all versions of the same view: Rule Contractualism.¹⁰⁴

1-2. The structure of Rule Contractualism

These views share a common structure. They all evaluate acts on the basis of the justifiability of the rules that permit them, where the justifiability of a rule is determined by what would happen if this rule was universally (or generally) adopted. As such, they are all versions of what I have called Rule Contractualism.

Just like RC, Rule Contractualism asks us to consider what would happen in an ideal world in which everyone (or almost everyone) accepted a set of rules which permits certain acts. The difference is that, instead of rules the universal acceptance of which would make things go best, Rule Contractualism prescribes following rules the universal acceptance of which would be appropriately justifiable to everyone. This means that, whereas the consequences of the general adoption of a set of rules matter intrinsically to rule consequentialists, they matter to rule contractualists only to the extent that they affect the justifiability of these rules, or to the extent that they provide reasons for acceptance or rejection. Instead of being concerned with maximising value, like RC, Rule Contractualism is thus concerned with a kind of hypothetical agreement. (I will discuss the implications of the similarities and differences between RC and Rule Contractualism in the next part.)

¹⁰⁴ For other examples of Rule Contractualist views, see Darwall (2006) p.301; Watson (1998) pp.174-176; and Nagel (1991), pp.36-37.

1-3. Interpretations of the requirement of justifiability

We have seen in chapter 1 that, for consequentialists, permissions and obligations are a function of the evaluative ranking of options according to a given axiology, and that changes in the axiology result in different moral views, some more plausible than others. In the same way, contractualists can construct a wide variety of moral views by amending their theories in two respects.¹⁰⁵

First, contractualists can interpret the requirement of appropriate justifiability differently. For instance, for Scanlon, appropriate justifiability means being reasonably unrejectable; for Southwood, it consists in being the object of hypothetical agreement between perfectly rational agents; and, for Parfit, it is determined by what everyone could rationally will. These nuances in contractualist views lead to different results.

As an illustration, let us look at the distinction between rationality and reasonableness. As pointed out by Scanlon, rationality is often taken to mean instrumental rationality, where the most rational thing to do is what best fulfils the agent's aim.¹⁰⁶ By contrast, reasonableness suggests the necessity of taking others' interests into account. So, whether one takes the subjects of justification to be rational or reasonable will lead to different rules being selected as the appropriately justifiable ones, and thus as the ones to be followed.

Consider the following example:

¹⁰⁵ See Suikannen (2020) on the structural differences between Contractualism and Consequentialism.

¹⁰⁶ Scanlon (1998) pp.191-192.

Water Rights: Two landowners are negotiating water rights. One of them, A, controls most of the water and does not need to cooperate, unlike the other landowner, B. Moreover, A is irritable and does not like her position being questioned.¹⁰⁷

Given the irritability of A, it would be rational for B to keep her requests to a minimum in order not to upset A, which might result in a situation which would be worse for B. It might also be rational for A to negotiate an agreement which guarantees that she will keep control of more water than B. However, it would be *unreasonable* for A to do so. We can also say that it would not be unreasonable for B to insist that they should share the water equally.

Therefore, if we interpret justifiability as what perfectly rational agents would agree to, as Southwood does, we might consider that a rule which would allow A to keep control of more water than B in situations such as the one in *Water Rights* is appropriately justifiable.¹⁰⁸ On the other hand, if we interpret justifiability as what cannot be reasonably rejected, as Scanlon does, it is probable that the rule which would allow A to keep more water than B would not be appropriately justifiable. Rather, the appropriately justifiable rule – that is, the one that cannot be *reasonably* rejected – might be the one which requires A to give B equal control over the water.

1-4. Theory of reasons or rationality

Second, once one interpretation of justifiability is selected, changes in the theory of reasons or rationality which ground the normative ranking of options can further alter the

¹⁰⁷ This example is from Scanlon 1998, p.192.

¹⁰⁸ This does not mean that Deliberative Contractualism would have this implication. Southwood (2010, p.99) seems to imply that, if A was perfectly deliberatively rational, she would be moved by B's "inviting her to consider the good she might do" by sharing the water. I am not sure this means it would be *irrational* for A to refuse to share. In any case, this is a great example of the way in which amendments to the theories of reasons or rationality supporting contractualist views can lead to different implications, as I explain in the next section.

implications of a rule contractualist view. For instance, deliberative contractualists might disagree on exactly what it is that perfectly deliberatively rational agents would agree on, and Scanlonian contractualists might disagree on what constitute reasons for reasonable rejection. The theory of rationality or reasons they adopt will significantly alter the first order moral view they obtain.

As an illustration, consider Scanlonian Contractualism on the issue of aggregation. Scanlon argues that reasons for rejection should be limited to personal reasons – they should concern facts about what would happen to an individual if a rule was generally accepted.¹⁰⁹ They should not be about what would happen to members of a group she is part of, or about what would happen *in general*. Scanlon refers to this as the Individualist Restriction. Scanlonian Contractualism, coupled with the Individualist Restriction, takes the following form: a rule cannot be reasonably rejected if it minimises the strongest individual complaint – that is, if reasonable objections to alternative rules are stronger.¹¹⁰ Since the Individualist Restriction entails that only claims made by individuals count as reasons to accept or reject a rule, whereas impersonal states of affairs cannot count as such reasons, in many cases, optimific rules can be reasonably rejected.

Take the following example:

Transmitter Room: Jones works in the transmitter room of a television station. There has been an accident in the middle of a World Cup match watched by millions of people, and a piece of equipment has fallen on his arm. He is in a lot of pain. To

¹⁰⁹ Scanlon (1998), p.229.

¹¹⁰ Following Parfit, Scanlon (1998, p.229) calls this the Complaint Model.

rescue him, we would have to turn off the transmitter, thereby interrupting the transmission of the game for fifteen minutes.¹¹¹

Here, rescuing Jones means imposing a minor inconvenience on millions of television viewers. Taken together, the cost imposed on these people is likely to outweigh the pain Jones will suffer. Therefore, the optimific rule – the one which would make things go best on aggregate – is “do not interrupt the broadcasting of the game”. However, according to the version of contractualism defended by Scanlon, we should only weigh against Jones’s claim to be rescued the individual claims of television viewers, and not their aggregate. Since none of their complaints against the rule “interrupt the broadcasting of the game to rescue Jones” individually outweighs Jones’s complaint against the principle “do not interrupt the broadcasting of the game”, the former is the rule which cannot be reasonably rejected.

In this example, we see that, if we adopt the Individualist Restriction, the optimific rule can be reasonably rejected. If we do not adopt the Individualist Restriction, however, and if we consider that aggregate wellbeing counts as a reason for rejection, optimific rules cannot be so rejected. Importantly, not all contractualists are as eager as Scanlon to reject aggregative conclusions.¹¹² For instance, Parfit contends that the rules that everyone can rationally will are also the rules that make things go best.¹¹³ He rejects the Individualist Restriction and argues that, since the goodness of a state of affairs gives us the strongest impartial reasons to want to bring it about, optimific rules are reasonably unrejectable. So, on Parfit’s view, unlike on Scanlon’s view, Rule Contractualism is extensionally equivalent to

¹¹¹ This example is from Scanlon (1998), pp.235-236.

¹¹² Moreover, it is unclear to what extent Rule Contractualism can avoid aggregative conclusions. Otsuka (2000) argues that Scanlon cannot avoid taking into account the claims of groups of individuals if he wants his view to have plausible partially aggregative implications (see Scanlon 1998, pp.238-241).

¹¹³ Parfit (2011), pp.398-401.

RC.¹¹⁴ This shows that the theories of reasons which ground rule contractualist views significantly affect their implications.

2. Rule Contractualism and RC

2-1. Similarities: Rule Contractualism, RC and the objections to AC

We have seen that rule contractualists disagree about what it takes for a rule to be justifiable – for instance, they disagree about what it is reasonable to accept and reject, or about what ideally rational agents would agree on. However, it seems very plausible that at least many of the rules that are justifiable to everyone, or that no one can reasonably reject, will also be the rules that make things go best. This is because, if a rule does not make things go best, that seems to give people good grounds to reject it.

Given this, and given the similarities in the structures of Rule Contractualism and RC, we can expect at least many versions of Rule Contractualism to have implications that are very similar to RC. This suggests that Rule Contractualism will be able to address the objections to AC that I discussed in chapter 1 in a way very similar to RC.

Take, for instance, the objection to AC from rights and duties. Consider again the example that we looked at in chapters 1 and 2:

Surgeon: a surgeon has five patients who will die if they do not get an organ transplant soon. A sixth patient is booked in for a benign surgery. During her

¹¹⁴ For other defences of the extensional equivalence between Rule Contractualism and RC, see Harsanyi (1955), Brandt (1979) and Levy (2013).

operation, the surgeon kills this patient and takes from her the organs necessary to save the five other patients, who live as a result.¹¹⁵

Remember that, according to AC, the surgeon's act is permissible, and even required. This is because this act would maximise good consequences, as five persons would survive instead of one. This conclusion is, of course, highly counter-intuitive.

According to RC, it is impermissible to kill one's patient to donate her organs. This is because, if everyone accepted a rule that permits doing so, no one would trust doctors anymore. People would stop seeking medical attention, and this would be a worse state of affairs than one in which people could still trust their doctors. Even though one instance of killing a patient to save five would have better consequences than letting them die, the cumulative effect of the internalisation, publicity and application of a rule that allows such acts would have worse consequences than that of a rule that forbids them. Therefore, the surgeon ought to act according to a rule which tells her not to kill her patient to donate her organs.

Rule Contractualism has similar implications, for similar reasons. The state of affairs that would result from the general acceptance of a rule permitting doctors to kill their patients to donate their organs is unjustifiable – it provides ground for reasonably rejecting this rule. This holds whether we believe that general states of affairs constitute reasons for rejection or not, since there would also be strong individual reasons to reject this principle.¹¹⁶

¹¹⁵ This example is from Foot (1978) and is discussed at length by Thomson (1985).

¹¹⁶ Rule contractualists who accept the Individualist Restriction have an added reason to consider a rule allowing doctors to kill their patients to donate their organs unjustifiable. Since the complaint of someone who gets killed by their doctor is bigger than the complaint of someone who dies from natural causes, any rule permitting to kill one's patient to save someone else can be reasonably rejected. Note that this rejection does not rely on universalisation, as even one such act can be reasonably rejected. However, rule contractualists do not unanimously adopt the Individualist Restriction. This is why I focus on the way in which the general structure of Rule Contractualism allows them to avoid the objections to AC.

Consider next the objection to AC regarding the way it deals with cases of collective impact. Consider again the example that we looked at in chapters 1 and 2:

Coffee: You see a documentary about the horrors of the coffee industry, including the exploitation of farmers and the use of slave and child labour. The next time you order a coffee, you ask for a Fairtrade brand. However, you find that the Fairtrade coffee is not as good as the one you usually buy, which is not Fairtrade. So, the next time you get a coffee, you order your usual brand.¹¹⁷

Remember that, according to AC, you do nothing wrong when you decide to purchase the coffee made under terrible working conditions. Since the impact of your single purchase is negligible, the satisfaction you will get from drinking the coffee whose taste you prefer is all that matters to your decision – it determines that buying the non-Fairtrade coffee results in a better state of affairs than buying the Fairtrade coffee. However, this result strikes most of us as mistaken: it seems impermissible to buy a cup of coffee which you know has been produced under miserable working conditions.

Collective action cases like *Coffee* are such that it is only when we look at the cumulative consequences of many individual actions of a certain kind that we can grasp their negative consequences. RC correctly implies that it is impermissible to buy the non-Fairtrade coffee, because it looks at the state of affairs that would result from the general acceptance of rules rather than at the consequences of a single act. Since the consequences of the general acceptance of a rule permitting us to buy non-Fairtrade coffee are much worse than that of a rule forbidding it, one ought to act in accordance with the latter.

¹¹⁷ This example is from Nefsky (2015).

Rule Contractualism gets the same result, from the same reasons. Since the state of affairs that would result from the general acceptance of a rule permitting us to buy non-Fairtrade coffee is very bad, it is also unjustifiable, and it provides everyone with grounds for reasonable rejection. This holds even if we accept the Individualist Restriction: the state of affairs that results from general acceptance of a rule permitting us to buy non-Fairtrade coffee is such that the farmers have strong *individual* objections to this principle.

2-2. Differences: the contractualist account of morality

Given the similarities between RC and Rule Contractualism, you might be wondering what reason there is to prefer one of these views over the other. The contractualist commitment to justifiability (rather than to making things go best) is a key reason why some favour Rule Contractualism over RC. Rule contractualists thus believe that their theory is appealing because we have good reason to want to act in ways that can be justified to others. For instance, Scanlon argues that, by acting in ways which are appropriately justifiable to others, one acknowledges them as rational and reasonable equals to whom justification is due, which in turn promotes valuable relationships of mutual recognition.¹¹⁸

The mere presence of a rule contractualist rationale is an advantage over RC. Recall from chapter 2 that RC seems to be lacking an explanatory rationale – one that justifies all of its implications. Unlike RC, Rule Contractualism provides us with a straightforward and appealing way to justify its conclusions.¹¹⁹ But this justificatory work does not stop there. If we accept the rule contractualist account of morality, the contractualist rationale I have presented also vindicates morality at large. This is why contractualists believe that their

¹¹⁸ Scanlon (1998), pp.158-168.

¹¹⁹ Note that even rule consequentialists seem to agree with the idea that the ideal of justifiability is appealing, since they attempt to provide a contractualist rationale for their view.

theory provides an account not only of right and wrong, but also of why we should care about right and wrong. It provides an account of the reasons we have to think about morality and to want to act in accordance with it.

This account has at least three advantages. First, it leaves space for first-order moral reasons alongside the higher-order reasons provided by the ideal of justifiability to others.¹²⁰ This is important in three respects. First, Rule Contractualism transcribes the complexity of moral reasoning. It does not negate the plurality of moral reasons by over-simplifying moral motivation, in the way consequentialist theories might do. On the contrary, it accommodates a plurality of moral reasons grounded in a plurality of values. Second, Rule Contractualism acknowledges the central role reasons play in morality by formalising the weighing of different reasons in a way reminiscent of pre-theoretical moral reasoning. Third, it seems phenomenologically accurate that, when we think about why we ought to act as morality demands, we often provide a plurality of reasons such as “I promised” or “he needs my help”, rather than a single one (such as “it would make things go best”).

This brings us to the second advantage of the rule contractualist account of motivation. Despite what I have just said about the presence of first-order reasons to act as morality demands, it seems just as phenomenologically accurate that we can also be moved by *higher-order* reasons to be moral of the kind rule contractualists suggest. Although we rarely go beyond the first-order reasons of the kind I mentioned when we engage in day-to-day moral thinking, we sometimes need a further justification for acting morally – one capable of motivating us to act in the appropriate way. The ideal of justifiability to others seems appealing enough to fulfil this function.¹²¹

¹²⁰ Ibid., pp.155-158.

¹²¹ See Pettit (2000) on concrete ways in which we might value standing in relationships of justifiability to others, and Lenman (2006) on instrumental reasons we have to want it.

Lastly, not only is the contractualist account of moral motivation phenomenologically accurate, but it also seems to give the right answer to what Scanlon calls “Prichard’s dilemma”.¹²² The dilemma is best understood as a dual requirement placed on any account of morality: when asked “why be moral?”, we need an answer which is not formulated in terms of self-interest – as such a reason to be moral would constitute the wrong kind of motivation – but we also need an answer which does not presuppose the normative pull of morality. The contractualist account of morality fulfils both criteria “by describing an ideal of relations with others which is clearly connected with the content of morality and, at the same time, has strong appeal when viewed apart from moral requirements”.¹²³

3. Does Rule Contractualism avoid the IWO?

3-1. The objection

The contractualist commitment to justifiability is a key reason why some favour Rule Contractualism over RC. Yet, despite this important difference with RC, Rule Contractualism is also vulnerable to the IWO. More specifically, if Rule Contractualism is similar enough to RC to avoid the objections to AC in a similar way, then Rule Contractualism is also similar enough to RC to be vulnerable to the IWO. (I will from here on assume that Rule Contractualism is able to avoid the objections to AC in a similar way to RC.)

Rule Contractualism has a two-level structure very similar to that of RC. Both assess acts indirectly, via rules whose selection relies on a process of universalisation; an act is

¹²² Scanlon (1998), pp.149-150.

¹²³ Ibid., p.155.

permissible if it conforms with a given set of rules which is selected on the basis of the consequences of its universal (or general) acceptance. Even though Rule Contractualism prescribes that we follow the rules that are justifiable, whereas RC prescribes that we follow the rules that make things go best, the test for permissibility remains basically the same: when deciding how to act, or when evaluating acts, we are asked to abstract from real-world circumstances.

This means that Rule Contractualism has absurd implications, just like RC, and for the same reasons. Consider, for instance, our example from chapter 2:

Boat: You are on a boat with nine other passengers. Nearby, another boat capsized, and there are now nine people in the water about to drown. There are nine lifejackets on your boat, which you and the other passengers could throw overboard to save the nine people from drowning. However, the lifejackets are stored in individual compartments with capricious locks, and opening them is slightly painful and annoying. Therefore, none of the other passengers of your boat bother doing anything. What should you do?¹²⁴

Most of us would agree that you should open all the compartments to throw nine lifejackets to the people drowning, since none of the other passengers in your boat are doing anything. However, just like RC, Rule Contractualism is unable to accommodate this intuition. This is because Rule Contractualism selects rules to follow on the basis of what would happen if *everyone* (or almost everyone) accepted them. If we look at what would happen if most people (say, 90 percent) accepted the rule “throw one lifejacket overboard” (or

¹²⁴ This case is inspired by Hooker (2002), p.164.

alternatively a more general rule like “contribute your fair share of efforts to save everyone”), we see that this rule would be appropriately justifiable. In such circumstances of quasi-total acceptance, no one would have reasonable ground to reject this rule, since every drowning person would be saved. However, following this rule in the real world, in which none of the other passengers are willing to throw lifejackets overboard, means letting eight persons drown. Yet, Rule Contractualism tells you to do just that: follow rules which would be appropriately justifiable if they were generally accepted. Therefore, Rule Contractualism might require you to act in ways that lead to disaster, such as the death of eight people in *Boat*, simply because such acts are permitted by appropriately justifiable rules.

This example highlights the problematic structural implications of evaluating acts indirectly, via rules which are themselves selected on the basis of what would happen at idealised levels of acceptance.¹²⁵ The problem is that, according to Rule Contractualism, just like with RC, what happens in the real world has virtually no influence on our evaluation of acts. This leads to implications which are not only intuitively implausible, but which also seem *unjustifiable*, at least in the sense that they could not be justified to people in the actual world, where there is less than complete acceptance of the ideal rules.

3-2. Mitigating the objection: flexible rules

The IWO was originally formulated against RC.¹²⁶ In chapter 2, I have presented a rule consequentialist response to this objection. We have seen that rule consequentialists like Hooker have claimed that RC does not lead to disasters, because the best set of rules

¹²⁵ Note that this objection applies whether we look at universal acceptance or general acceptance. This is why it is not enough to provide a convincing account of the reason why a justifiable code would contain rules to deal with non-compliance and moral education, as Suikkanen (2014) does.

¹²⁶ See Parfit (2011) pp.312-320, and Podgorski (2018).

includes an overriding Avoid Disaster rule. However, I have argued that it is not the case that the set of rules which would have the best consequences if it was generally adopted would contain an Avoid Disaster rule. Moreover, I have shown that, even if we accept this clause, RC nevertheless entails the counterintuitive implication that one should bring about bad-although-not-disastrous consequences, since rule consequentialists resist the collapse of their theory into AC.

The same reasoning applies to Rule Contractualism. In response to the IWO, rule contractualists could argue that, since the most justifiable set of rules includes an overriding Avoid Disaster rule, their theory does not entail that we should act in ways that lead to disaster in cases such as *Boat*. But this will not do. Although an Avoid Disaster rule makes Rule Contractualism less implausible in cases such as *Boat*, it still entails that we ought to act in ways that remain highly counter-intuitive, even though they do not lead to disasters. Imagine a modified version of *Boat*, in which the nine people in the water risk blindness instead of death, or any harm you think would be short of a disaster: Rule Contractualism would then still require that you follow the rule which is justifiable when it is generally accepted, and thus that you only throw one lifejacket overboard. Again, it seems like this implication is not only counter-intuitive, but also that it could be reasonably rejected in the real world by the nine who are at risk of harm, even if it is not considered to be a disaster.

One solution open to rule consequentialists is to take the Avoid Disaster provision even further. Scanlon, for instance, suggests that his view does not have such problematic implications, and that there is no tension between Rule Contractualism's dictates and real-world justifiability. He argues that principles are flexible and require interpretation. It is "normally" the case that we ought to follow them, and "normally" covers many

qualifications.¹²⁷ Taking the example of promises, Scanlon argues that any principle which does not start with “in the absence of special justification” could be reasonably rejected.¹²⁸ From his discussion of the use of principles in his theory, I take it that similar qualifications apply to rules covering other areas of morality as well. Such provisions mean that, when there are strong reasons to go against a justifiable rule, or when the reasons to accept a justifiable rule are weakened, one is permitted to act differently from what the rule would prescribe. It ensures that following justifiable principles does not lead to counter-intuitive results such as in our modified *Boat* example, and that it does not mean acting in ways that cannot be justified to people in the actual world. Does this response address the IWO?

3-3. Collapse?

First, let me note that it is not obviously the case that a principle which does not start with “in the absence of special justification” could be reasonably rejected in ideal circumstances of general acceptance. Any principle which would be justifiable in ideal circumstances could already include clauses to address situations which occur at this idealised acceptance level. This means that, even without the qualification “in the absence of special justification”, these principles could not be reasonably rejected. In other words, as long as a principle is specific enough to ensure that it could not be reasonably rejected in all the situations which occur at idealised acceptance levels, it does not need to start with “in the absence of special justification” to be reasonably unrejectable.

Think, for instance, about a world in which 90% of the population accepted the rule “throw one lifejacket overboard”. In such a world, none of the drowning people in *Boat* could

¹²⁷ Scanlon (1998) pp.197-202.

¹²⁸ Ibid. pp.298-299.

reasonably reject this rule, even if it does not start with “in the absence of special justification”, because all of them would be saved anyway.¹²⁹ Even if we complicate the world of this thought experiment to make it more realistic, such that there would be non-standard cases in which someone could reasonably object to this simplistic rule, we can modify the rule so that it would include a clause which addresses such situations. In principle, this can be done indefinitely for all the situations which would occur at a given acceptance level, so that there would be no need to add “in the absence of special justification” for this rule to be reasonably unrejectable. But this still does not mean that this unrejectable rule would be qualified in the way Scanlon hopes – that is, it does not mean that it would permit exceptions in the real world, when the situation that requires an exception would not present itself in ideal conditions of general acceptance.

Scanlon might respond that a principle which would be specific enough to be unrejectable in all possible situations would be too complex to ever formulate, and that there would always be nonstandard cases that are not covered by any formulable principle, even in an ideal world. After all, the ideal world we consider when we want to determine which rules are justifiable only differs from the real world with respect to the acceptance level of a given set of rules. Apart from that, it is in every respect similar to the real world, including in its complexity. Therefore, it is likely that, even in an ideal world of general acceptance, there would be someone who could reasonably reject an unqualified principle – one which does not start with “in the absence of special justification”.¹³⁰

¹²⁹ One could object that this rule is too specific, and that it is not what Scanlon has in mind when he defends the use of principles (see Scanlon 1998, pp.197-202). However, the same reasoning applies to general rules such as “contribute your fair share of efforts to save everyone”. Even if this rule does not start with “in the absence of special justification”, no one could reasonably reject it in a world in which everyone contributed fairly, because everyone would be saved anyway.

¹³⁰ This allows Rule Contractualism to avoid my reformulation of the IWO against RC, which was grounded in the fact that the Avoid Disaster rule would *not* be part of the best set of rules. Cases in which there exist special justifications not to follow a rule seem much more numerous than cases in which following a rule would lead to disaster. Therefore, although formulable and internalisable rules might be specific enough to avoid disasters, they cannot be specific enough to cover all potential special

Now, if we accept that unrejectable principles start with “in the absence of special justification”, it seems like Rule Contractualism might indeed be able to address the IWO. In *Boat*, for instance, the fact that eight people will drown (or be harmed, in the modified *Boat* example) constitutes special justification not to follow a principle which tells us to throw only one life jacket overboard, or to only contribute one’s fair share of efforts. If this is the case, Rule Contractualism does not have the counter-intuitive implication that you ought to let eight persons drown or be harmed. At least in this example, it does not entail that you ought to act in ways that can be reasonably rejected in the actual world for the sake of acting according to a rule which would be justifiable if it was generally accepted.

However, this does not mean that there will not be cases in which Rule Contractualism prescribes acting in ways which can be reasonably rejected in the actual world. Rule contractualists must specify what constitute sufficient reasons not to follow a principle. In *Boat*, it seems quite uncontroversial that the fact that eight people would be harmed (short of disaster) if you only contributed your fair share of efforts to help those in need gives you strong reasons to deviate from the original principle. But what about cases in which the consequences of following a principle are less catastrophic? Should we deviate from the rule every time following it can be reasonably rejected in the real world– that is, every time there is a tension between ideal justifiability and real-world justifiability? If the answer is no, Rule Contractualism will sometimes entail acting in ways which can be reasonably rejected by people in the real world. If the answer is yes, Rule Contractualism collapses into another view, Act Contractualism, which I will present in the next section.

justifications. This makes the claim that appropriately justifiable rules should contain an “in the absence of special justification” clause more plausible than Hooker’s claim that the best code contains an Avoid Disaster rule.

In a nutshell, the idea is that, if rule contractualists accept that one should deviate from justifiable principles every time someone can reasonably object to the application of these principles in the real world, then their theory simply prescribes acting in ways that are appropriately justifiable every single time. In other words, Rule Contractualism collapses into an act-based version of contractualism, in much the same way RC is accused of collapsing into AC. I will argue that this collapse is not that bad an option after all.

4. Rule Contractualism and Act Contractualism

I argued in chapter 2 that, if rule consequentialists want to address the IWO, they have to accept the collapse of their theory into AC. I have just argued that, in much the same way, Rule Contractualism could collapse into an act-based version of that view. According to this act-based version, which we will call

Act Contractualism: an act is permissible if and only if it is appropriately justifiable to everyone.

Act Contractualism, like AC, evaluates acts directly, instead of evaluating them for their conformity to rules. Unlike AC, however, it is concerned with appropriate justifiability, rather than with making things go best. Just as in the case of Rule Contractualism, what constitutes appropriate justifiability can be interpreted in different ways, leading to the formulation of different first-order theories. Depending on the account of appropriate justifiability which is selected, and on the theory of reasons or rationality that grounds it, versions of Act

Contractualism can be more or less plausible, and be vulnerable or immune to different objections.

In this section, I ignore the differences between different versions of Act Contractualism to focus on the structural differences between this view and Rule Contractualism. Some of these differences correspond to reasons rule contractualists have for resisting the collapse of their view into Act Contractualism, whereas others are reasons to accept it.

4-1. Reason to resist the collapse: avoid the objections to AC

Presumably, rule contractualists, just like rule consequentialists, want to avoid the collapse of their theory into an act-based version of it. Scanlon, for instance, suggests that we ought to deviate from a justifiable principle only when there are *strong* reasons to do so.¹³¹ This suggests that, whenever the reasons to object to the application of an appropriately justifiable principle in the real world are less than strong, Scanlonian Contractualism will still prescribe that we follow this principle, even if doing so can be reasonably rejected by someone in the real world. But why not act in ways that are justifiable in the real world – that is, why not accept the collapse into Act Contractualism?

If rule contractualists accept the collapse of their theory into Act Contractualism, it seems like it becomes vulnerable to the objections against AC that I have presented in chapter 1. For example, remember the objection from collective impact, and think again about *Coffee*. Given that your act of buying non-Fairtrade coffee, when taken in isolation, has almost no impact, there is no one to object to your single act of buying non-Fairtrade coffee, whereas you could reasonably object to being unable to buy the coffee you like the most. If Scanlonian contractualists granted that one can permissibly deviate from a justifiable principle every time

¹³¹ See Scanlon (1998), pp.199-201 and 298-299.

someone can reasonably object to its application in the real world, it seems like they would have to accept that, in *Coffee*, not being able to buy the coffee you like the most constitutes a strong enough reason to object to the principle which tells you to buy Fairtrade coffee. Therefore, they would have to accept that, in the real world, you ought to buy non-Fairtrade coffee, which is the counter-intuitive conclusion that grounded the objection from collective impact against AC.¹³²

In a similar way, Rule contractualists risk losing their defence against other objections to AC if they accept the collapse of their theory into Act Contractualism. Rule Contractualism would imply that you ought to deviate from principles permitting reasonable degrees of self-concern or of partiality to one's loved ones, or from principles forbidding lying and stealing, as soon as someone could reasonably object to the application of an unrejectable principle in the real world – even if their objections are weak.

I believe that some versions of Act Contractualism might be able to address at least some of these objections.¹³³ However, whether they can will depend on the theory of reasons or rationality that is adopted. An Act-based version of contractualism loses the benefit of addressing the objections to AC structurally, in the way RC and Rule Contractualism do.

4-2. Reason to accept the collapse: justifiability in the real world

Accepting that one should deviate from an unrejectable principle every time someone can reasonably object to the principle being applied in the real world amounts to adopting Act Contractualism. This means losing the advantage of a two-level structure, which allowed Rule

¹³² This applies even to versions of Rule Contractualism which reject the Individualist Restriction. Since the consequences of a single act of buying a non-Fairtrade coffee are close to (if not) null, the badness of the resulting state of affairs does not constitute a reason not to buy non-Fairtrade coffee.

¹³³ For reasons of space, this will have to be the topic of future research.

Contractualism to avoid the objections to AC in a similar way to RC. However, resisting this collapse and placing the bar for what constitutes a strong enough justification to deviate from a justifiable principle any higher also comes at a cost. It implies that following Rule Contractualism will sometimes require acting in ways that can be reasonably rejected in the real world.

Think, again, about our modified *Boat* example. There is a threshold below which the harm faced by the nine people in the water is not important enough to constitute “strong” justifications to deviate from a rule that prescribes to contribute only one’s fair share of efforts to help others. Since this principle is appropriately justifiable as a general standard of behaviour – that is, since its general acceptance would be appropriately justifiable – and since the “in the absence of special justification” provision does not apply, Rule Contractualism then entails that you must throw only one life jacket overboard. Yet, it seems like the nine persons in the water have reasonable claims to be rescued, as long as the harm they face outweighs the inconvenience you would endure because of the capricious locks.

Rule contractualists place no bar on how strong a reason to object to a principle must be when they determine which principles are appropriately justifiable. For instance, nothing says that a claim not to suffer a minor harm does not provide reasonable grounds for rejecting a principle, when the complaints against alternate principles are even less important. Yet, rule contractualists, if they are to resist the collapse of Rule Contractualism into Act Contractualism, place such a limit when they determine when we can permissibly deviate from a justifiable principle. This is odd: why wouldn’t the same claim constitute reasonable grounds to object to the application of a principle in the real world, when it is a reason to object to a principle as a public standard of behaviour?

In the spirit of the Inconsistency Objection levelled against RC, one might be tempted to say that this oddity makes Rule Contractualism inconsistent, because Rule Contractualism does not prescribe that we act in ways that are appropriately justifiable in the real world. However, this need not be the case. For Scanlon and other rule contractualists, appropriate justifiability might consist in just this: conformity with rules which would be appropriately justifiable if they were generally followed.

Nevertheless, there is a sense in which this rule contractualist account of justifiability is unsatisfying. As Hanoch Sheinman points out, it is not clear exactly what is attractive in a relation of conformity to justifiable rules.¹³⁴ He thus argues that Rule Contractualism derives its appeal from the idea that it allows one to act in ways that are justifiable to others, but that this appeal is ill-deserved.

This seems right. Recall the motivation to adopt Rule Contractualism which I have presented in section 2: the reasons we have to adopt Rule Contractualism, Scanlon tells us, are grounded in “the reasons we have to live with others on terms that they could not reasonably reject”.¹³⁵ However, I have just shown that Rule Contractualism sometimes requires us to act in ways which others can reasonably reject. Therefore, the ideal of acknowledging others as reasonable equals to whom justification is due, which is supposed to derive from acting in ways others cannot reasonably object to, do not seem to be fulfilled by Rule Contractualism. Quite the contrary: it seems plausible that treating others as reasonable equals to whom justification is due entails acting in ways which they cannot reject *in the real world*. Therefore, the compelling reasons rule contractualists offer in favour of their

¹³⁴ Sheinman (2011).

¹³⁵ Scanlon (1998), p.154.

view actually supports Act Contractualism and its conception of justifiability as *real-world* justifiability.

Although accepting the collapse into Act Contractualism means giving up on the two-level structure that made Rule Contractualism immune to some of the main objections against AC, it seems like the contractualist ideal of treating others as reasonable equals might be better fulfilled by Act Contractualism. Given this, and given that some versions of Act Contractualism might be able to address some of the objections against AC, rule contractualists have good reason to accept the collapse of their view into Act Contractualism.

5. Act Contractualism and AC

In the previous chapter, I argued that, because of the IWO, rule consequentialists have good reason to retreat back to AC, despite the problems this view faces. In this chapter, I have argued that the ideal of living with others in terms they could not reasonably reject gives us good reason to accept Rule Contractualism rather than RC. I have also argued that, since this ideal might in fact be better fulfilled by Act Contractualism than by Rule Contractualism, we should accept the collapse of Rule Contractualism into Act Contractualism. I will now briefly compare Act Contractualism and AC.

Act Contractualism and AC have a similar structure; both assess acts directly. The difference is that Act Contractualism cares about appropriate justifiability, whereas AC cares about good consequences. Therefore, those who were convinced by the contractualist rationale I have presented – those who care about justifiability and preferred Rule Contractualism to RC – should prefer Act Contractualism to AC. Since these views have different foundations, Act Contractualism would not be redundant even if it turned out to have

the exact same implications as AC, which is a good reason to take both views seriously and consider them separately.

However, we have good reason to think that Act Contractualism and AC have different implications. This is the case, for instance, regarding the ethics of future people. Consider the problem known as

The Procreation Asymmetry:

- 1) It is impermissible to bring into existence people with lives less than worth living, other things being equal.
- 2) Yet, it is permissible not to bring people with lives worth living into existence, other things equal.¹³⁶

The *Procreation Asymmetry* captures our intuitions that, although it would be wrong for a couple to bring into existence a child whom they know would have an existence not worth living, they do not have a duty to bring into existence a child whom they know will have a life worth living.

AC seems unable to capture this asymmetry. It can capture claim 1, since it is plausible that a state of affairs in which someone leads an existence less than worth living is worse than a state of affairs in which no one does. The problem is that, by the same logic, AC implies that we *do* have a duty to bring into existence people with lives worth living. This is because a state of affairs in which someone lives a life worth living is better than a state of affairs in which no one does. Therefore, AC cannot capture claim 2.

¹³⁶ For discussions of the Procreation Asymmetry, see McMahan (1981) and (2009), Elstein (2005) and Roberts (2011).

Rule Contractualism, on the other hand, captures the *Procreation Asymmetry*. We must act in ways that are appropriately justifiable to the people that exist; and bringing people with miserable lives into existence would be unjustifiable to them. However, when we fail to bring into existence people with lives worth living, there is no one to reasonably reject our (in)action. Therefore, Act Contractualism captures both claim 1 and claim 2.

There is a catch, however. This response to the procreation asymmetry suggests that Act Contractualism is unable to address what is known as

The Non-Identity Problem:

- 1) If a woman has a choice between bringing into existence a child with an impairment, or delaying her pregnancy to bring into existence a child without impairment, she ought to delay her pregnancy, even if the child with the impairment would still have a life worth living.
- 2) The child the woman would give birth to if she did not wait would be different from the child she would give birth to if she waited, given that they would be produced from different reproductive cells.¹³⁷

The *Non-Identity Problem* illustrates our intuition that we should make future generations as well-off as possible, other things equal, even if this requires taking actions which would affect the lives of individuals in such a way that they might conceive at different times than they would otherwise have.

Act Contractualism is vulnerable to the *Non-Identity Problem* because it cannot account for claim 1. Since the child born with an impairment would not exist if her mother had waited

¹³⁷ For discussion of the Non-Identity Problem, see Parfit (1986), pp.351-379 and Boonin (2014).

to conceive, she is not made worse-off if her mother does not delay her pregnancy. Since it would not be *her* who would be born without impairment if her mother had waited, but *another* child, she has no ground for complaint. In other words, since no one would be wronged if the woman did not delay her pregnancy, it is not impermissible for her not to.

On the other hand, AC does not face the *Non-Identity Problem*. Since things would go best if the woman delayed her pregnancy, it is impermissible for her not to, even if the child she gives birth to if she waits is not the one which would have otherwise been born.

To conclude, Act Contractualism and AC have different implications regarding future generations. While Act Contractualism captures the *Procreation Asymmetry*, AC does not; and while Act Contractualism is vulnerable to the *Non-Identity Problem*, AC is not. We thus have good reason to think that Act Contractualism and AC do not only have different foundations, but also that they have different implications, which gives us additional grounds to weigh them against each other. This seems to me a promising avenue for future research.

Conclusion

In this dissertation, I have carved out conceptual space for Act Contractualism. I have not argued that it is the best theory, but only that it deserves more attention than it has hitherto received. Since it rests on a distinct, appealing rationale, and since it has different implications than Rule Contractualism and AC, it should be taken seriously.

Questions remain regarding whether Act Contractualism avoids some of the objections raised against AC, and whether it is vulnerable to new objections. My hope is that these questions will be investigated, and that Act Contractualism will get attention in its own right. Although more research has to be done into what the best version of this view might be, it

seems to be a serious contender to the three other views I have considered as part of this project.

Bibliography

Anscombe, G. E. M. (2005). *Human Life, Action and Ethics: Essays*, Exeter: Imprint Academic

Ashford, E. (2000). "Utilitarianism, Integrity and Partiality", *The Journal of Philosophy*, 97(8), pp.421-439

Arneson, R. (2005). "Sophisticated Rule Consequentialism: Some Simple Objections", *Philosophical Issues*, 15(1), pp.235-251

Bentham, J. (1823). *The Principles of Morals and Legislation*, London: T. Payne

Boonin, D. (2014). *The Non-Identity problem and the ethics of future people*, Oxford: Oxford University Press

Brandt, R. (1979). *A Theory of the Good and the Right*, Oxford: Oxford University Press

Brandt, R. (1988). "Fairness to Indirect Optimific Theories in Ethics", *Ethics*, 98(2), pp.341–60

Brandt, R. (1989). "Morality and Its Critics", *American Philosophical Quarterly*, 26(2), pp.89-100

Brink, D. (1993). "The separateness of persons, distributive norms and moral theory" in Frey and Morris (eds.) *Value, Welfare and Morality*, Cambridge: Cambridge University Press, pp.252-289

Card, R. (2007). "Inconsistency and the Theoretical Commitments of Hooker's Rule-Consequentialism", *Utilitas*, 19(2), pp.243-258

Darwall, S. (2006). *The Second-Person Standpoint*, Cambridge MA: Harvard University Press

Dreier, J. (2011). "In Defence of Consequentialising" in Timmons (ed.) *Oxford Studies in Normative Ethics Volume 1*, pp.97–119

Elstein, D. (2005). "The Asymmetry of Creating and Not Creating Life", *Journal of Value Inquiry*, 39(1), pp.49–59

- Foot, P. (1978). "The Problem of Abortion and the Doctrine of the Double Effect" in *Virtues and Vices and Other Essays in Moral Philosophy*, pp.19-31
- Gert, B. (2005). *Morality: Its Nature and Justification*, Oxford: Oxford University Press
- Harman, E. (2015). "The irrelevance of moral uncertainty", in Shafer-Landau (ed.) *Oxford Studies in Metaethics Volume 10*, pp.53-78
- Harsanyi, J. (1955). "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy*, 63(4), pp.309–21.
- Hooker, B. (2002). *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*, Oxford: Oxford University Press
- Hooker, B. (2005). "Reply to Arneson and McIntyre", *Philosophical Issues*, 15(1), pp.264-281
- Hooker, B. (2007). "Rule-consequentialism and Internal Consistency: A Reply to Card", *Utilitas*, 19(4), pp.514-519
- Hooker, B. and Fletcher, G. (2008). "Variable versus Fixed-Rate Rule-Utilitarianism", *The Philosophical Quarterly*, 58(231), pp.344-352
- Horton, J. (2017). "Aggregation, Complaints, and Risk", *Philosophy & Public Affairs*, 45(1), pp. 54-81.
- Horton, J. (2018). "Always Aggregate", *Philosophy & Public Affairs*, 46(2), pp. 160-174
- Horton, J. (2020). "Aggregation, Risk and Reductio", *Ethics*, 130(4), pp.514-529
- Jackson, F. (1991). "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection", *Ethics*, 101(3), pp.461-482
- Kagan, S. (2011). "Do I Make a Difference?", *Philosophy & Public Affairs*, 39(2), pp.105-141
- Kamm, F. (1993). *Morality, Mortality, Volume 1*, Oxford: Oxford University Press
- Kupperman, J. (1981). "A Case for Consequentialism", *American Philosophical Quarterly*, 18(4), pp.305-313
- Lenman, J. (2006). "Compatibilism and contractualism: the possibility of moral responsibility", *Ethics*, 117(1), pp.7–31
- Levy, S. (2013). "A Contractualist Defence of Rule Consequentialism", *Journal of Philosophical Research*, 38, pp.189–201
- McCloskey, H. (1965). "A Non-utilitarian Approach to Punishment", *Inquiry*, 8(1–4), pp.249–263

- McMahan, J. (1981). "Problems of Population Choice", *Ethics*, 92(1), pp.96–127
- McMahan, J. (2009). "Asymmetries in the Morality of Causing People to Exist", in Roberts and Wasserman (eds.) *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, Dordrecht: Springer, pp.49–68
- Mill, J.S. (1871/1969). "Utilitarianism", in Robson (ed.) *The Collected Works of John Stuart Mill, Volume X: Essays on Ethics, Religion and Society*, Toronto: Toronto University Press, pp.205-259
- Nagel, T. (1970). *The Possibility of Altruism*, New Jersey: Princeton University Press
- Nagel, T. (1979). *Mortal Question*, Cambridge: Cambridge University Press
- Nagel, T. (1991). *Equality and Partiality*, Oxford: Oxford University Press
- Nesky, J. (2012). "Consequentialism and Collective Harm: A Reply to Kagan", *Philosophy and Public Affairs*, 39(4), pp.364-395
- Nesky, J. (2015). "Fairness, Participation, and the Real Problem of Collective Harm" in Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 5*, pp.245-270
- Otsuka, M. (2000). "Scanlon and the claims of the many versus the one", *Analysis*, 60(3), pp.288-293
- Parfit, D. (1986) *Reasons and Persons*, Oxford: Oxford University Press
- Parfit, D. (1997). "Equality and priority", *Ratio*, 10(3), pp.202-221
- Parfit, D. (2011). *On What Matters*, Oxford: Oxford University Press
- Pettit, P. (2000). "A Consequentialist Perspective on Contractualism", *Theoria*, 66(3), pp. 228–45
- Podgorski, A. (2018). "Wouldn't it be Nice? Moral Rules and Distant Worlds", *Noûs*, 52(2), pp. 279-294
- Portmore, D. (2011). *Commonsense consequentialism: Wherein morality meets rationality*, New York: Oxford University Press
- Railton, P. (1984). "Alienation, Consequentialism, and the Demands of Morality", *Philosophy and Public Affairs*, 13(2), pp.134-171
- Rachels, J. (1975). "Active and Passive Euthanasia", *The New England Journal of Medicine*, 292 (2), pp.78-80
- Rajczi, A. (2016). "On the Incoherence Objection to Rule-Utilitarianism", *Ethical Theory and Moral Practice*, 19(4), pp. 857-876

- Rawls, J. (1999). *A Theory of Justice*, Oxford: Oxford University Press
- Ridge, M. (2009). "Climb Every Mountain?", *Ratio*, 22(1), pp. 59-77
- Roberts, M. (2011). "An Asymmetry in the Ethics of Procreation", *Philosophy Compass*, 6(11), pp.765- 776
- Rosen, G. (2009). "Might Kantian Contractualism be the Supreme Principle of Morality?", *Ratio* 22(1), pp. 78–97
- Scanlon, T. M. (1998). *What We Owe to Each Other*, Cambridge, MA: The Belknap Press of Harvard University Press
- Sheinman, H. (2011). "Act and Principle Contractualism", *Utilitas*, 23(3), pp.288-315
- Singer, P. (1993). *Practical Ethics*, Cambridge: Cambridge University Press
- Smart, J. (1956). "Extreme and Restricted Utilitarianism", *The Philosophical Quarterly*, 6(25), pp. 344-354
- Smart, J. (1973). "An Outline of a System of Utilitarian Ethics" in Smart and Williams (eds.) *Utilitarianism: For and Against*, London: Cambridge University Press. pp. 1-70
- Smith, H. (2010). "Measuring the Consequences of Rules", *Utilitas*, 22(4), pp.413-433
- Southwood, N. (2010). *Contractualism and the foundations of morality*, Oxford: Oxford University Press.
- Southwood, N. (2019). "Contractualism for Us As We Are", *Philosophy and Phenomenological Research*, 99(3), pp 529-547
- Suikkanen, J. (2014). "Contractualism and the Conditional Fallacy", in Timmons (ed.) *Oxford Studies in Normative Ethics, Volume 4*, Oxford: Oxford University Press, pp.113-136
- Suikkanen, J. (2020). *Contractualism*, Cambridge: Cambridge University Press.
- Taurek, J. (1977). "Should the Numbers Count?", *Philosophy and Public Affairs*, 6(4), pp.293-316
- Thomson, J. J. (1985). "The Trolley Problem", *The Yale Law Journal*, 94(6), pp.1395-1415
- Tomlin, P. (2017). "On Limited Aggregation", *Philosophy and Public Affairs*, 45(3), pp.232-260
- Voorhoeve, A. (2014). "How Should We Aggregate Competing Claims?", *Ethics*, 125(1), pp.64-87

Voorhoeve, A. (2017). "Why One Should Count Only Claims with which One Can Sympathize", *Public Health Ethics*, 10(2), pp.148-156

Watson, G. (1998). "Some Considerations in Favor of Contractualism", in *Rational Commitment and Social Justice*, Cambridge: Cambridge University Press, pp. 168-185

Wiland, E. (2010). "The Incoherence Objection in Moral Theory", *Acta Analytica: Philosophy and Psychology*, 25(3), pp. 279-284

Williams, B. (1973). "A Critique of Utilitarianism" in Smart and Williams (eds.) *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, pp.77-135

Williams, B. (1981). "Persons, Character Morality" in *Moral Luck: Philosophical Papers 1973-1980*, Cambridge: Cambridge University Press, pp. 1-20