Moral Limits on Climate Policy

A Rule-emphasizing Utilitarian Perspective on Geoengineering

Lage Nøst

Master thesis in Philosophy

Supervisor: Prof. Arne Johan Vetlesen

UNIVERSITY OF OSLO

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities

Fall 2020

II

Lage Nøst

Moral Limits on Climate Policy

A Rule-emphasizing Utilitarian Perspective on Geoengineering

2020

Moral Limits on Climate Policy: A Rule-emphasizing Utilitarian Perspective on Geoengineering

Lage Nøst

Summary

Humanity's insufficient attempts at reducing carbon emissions to prevent runaway climate change have inspired some to propose that we "engineer" the climate-system; large-scale technological interventions, with the aim to manage solar radiation and remove carbon dioxide from the atmosphere, could possibly keep dangerous global warming at bay – at least as a last resort if conventional mitigation efforts fail. However, such "geoengineering" could also possibly involve violations of someone's moral rights. If so, should we do it anyway, despite rights infringement? Or should we refrain out of concern for the moral rights of affected groups or individuals, even if geoengineering could effectively prevent or minimize harmful effects of ongoing climate change? To pose the question more generally: Are there moral rights which could, in principle, block effective climate policies?

In this thesis, I attempt to answer these questions in two steps: First, I defend a theory of normative ethics. Then, I apply it to a "geoengineering vs. rights" scenario. Specifically, I defend a formulation of Utilitarianism which emphasizes moral rules, and discourages calculation of consequences under normal circumstances, but recommends being cognizant that general moral rules can justifiably be broken under exceptional circumstances if the principle of utility clearly demands it. This translates to the conclusions that rights claims could, in principle, be set aside by policymakers to prevent significantly greater harms or losses, and that emergency geoengineering can't be ruled out as unethical out of hand. I argue that – other than the overriding concern to maximize the sum total of net pleasure in the experienceable universe – there are no strict moral limits on climate policy, but that humans, given our inherent epistemic barriers, would be wise to downplay this fact and limit ourselves under most circumstances, for example by formulating and adhering to what I call "Rules (of thumb)" – heuristic devices, informed by the principle of utility, which could helpfully guide us in practical moral reasoning. I propose some Rules (of thumb) for policymakers considering geoengineering schemes.

VI

Acknowledgements

Table of contents

1 Introduction

Saving the world from catastrophic climate disruption seems very important. But how important? Assuming there is a true morality, would it require that we did *everything* in our power to stop runaway climate change? Would all policies be justified by default if they could avert catastrophe? Or could legitimate rights claims obstruct certain policies? Are some rights truly inviolable? In other words, are there moral limits on climate policy?

In what follows, I hope to answer – or begin to answer – some of these questions. They are, I believe, pressing. Climate change is now impacting communities and ecosystems around the world, and will hit us harder in decades to come (IPCC 2014). Already, policymakers have begun considering audacious countermeasures of unprecedented magnitude (and, some argue, unprecedented folly). Under the header "geoengineering", a wide array of plans for large-scale, deliberate intervention in the climate system are being seriously discussed (Royal Society 2009). Hopes are we could cool the Earth and save its inhabitants from the worst effects of human-induced global warming. Critics, however, argue that such "technofixes", if we could get them to work, would screw things up even worse, exacerbating extant injustices and tying us to the mast of the sinking ship of consumerist capitalism.

Soon, the shape of geoengineering discussions could shift from largely theoretical musings to real-world policy debates over testing and implementation at scale (cf. Boettcher and Schäfer 2017). If geoengineering is implemented, it will affect everybody – some of whom will be affected negatively, perhaps having their fundamental rights violated. How, then, ought decision-makers think about the ethics of geoengineering? Which principles ought to guide their deliberations concerning whether or not to go ahead with some climate policy, despite seemingly legitimate rights claims opposing it?

1.1.1 Thesis outline

This thesis has two parts. First, in "Part I", after setting the stage by discussing how rights should be properly conceived, I develop a theory of normative ethics to be applied in the second part. Drawing on, among others, Hare (1981), and his account of the *intuitive* and *critical* levels of moral thinking, I suggest what I take to be the most viable formulation of Utilitarianism, according to which moral rules should be emphasized and Utilitarian calculation should normally be discouraged. I argue, however, that exceptional circumstances might warrant our attempting to assess likely consequences of different options, weighing them in light of the Utilitarian criterion of right action, and that no rule-worshipping moral theory could handle exceptional circumstances successfully. I discuss epistemic barriers and other objections to this kind of thinking. I spend quite some time replying to objections against foundational Act Utilitarianism, seeing as much of what I say in "Part II" hinges on the plausibility of this view.

Then, in "Part II", I set out to apply my normative theory to the geoengineering case. I expound on different geoengineering schemes – their promise and risks – before I consider principles pertaining to practical reasoning in emergencies. I then explore whether, in a possible future scenario where geoengineering has become necessary to contain dangerous global warming, morality would require us to commence geoengineering, despite possible infringement on otherwise legitimate rights. I focus on a particular class of geoengineering methods, going by the moniker "solar radiation management", because such schemes are more likely to be pursued as candidates for emergency deployment (they could be ramped up quickly and would have a near-instant cooling effect on the planet). I discuss different rights that could stand against such interventions, assessing them in light of the criterion of rightness I identified in the first part, and suggest some helpful patterns of thinking for the benefit of policymakers facing geoengineering dilemmas. I briefly consider further objections to my main findings, namely that (1) no rights claim is ultimately above the consequences of adhering to it, which means (2) geoengineering cannot be ruled out on principle.

When defending some argument or drawing out some intuition in "Part I" I have not limited myself to rely on examples relevant to the geoengineering case. These arguments concern ethics in general, and I would rather put forth philosophical arguments in their strongest form, than necessarily design them to comport with the very questions I hope they can

elucidate. Hence, due to limited cross-referencing, "Part I" could for the most part be read as a standalone work. This goes for chapter 4 as well, which is exegetical, but not for the rest of my discussion on geoengineering in "Part II", seeing as it relies heavily on the normative framework I establish in "Part I".

1.1.2 A note on emergency cases

I take a keen interest in the ethics of emergencies. This thesis is motivated in no small part by the intuition that otherwise prohibited acts could be justified under exceptional circumstances. For example: Killing, though normally frowned upon, could be justified to save your family from some axe-murderer (and perhaps geoengineering could be justified to save life on Earth). I believe, like Peter Singer (1972) and others, that emergencies – while being situations out of the ordinary, often warranting extraordinary behavior to promote moral goods – are not special in terms of which fundamental moral principles apply (Sterri and Moen 2020). Rather, they can serve to make salient what our moral duty *actually* is. To leverage this effect, many of the imaginary cases I use to bolster my claims herein – my "intuition pumps", if you will (Dennett 2013) – are construed as emergency scenarios. Some philosophers claim that emergency cases should be assessed relative to a criterion of moral rightness *different* from that pertaining to circumstances of normalcy (ibid.), and that such thought experiments are therefore uninformative with regard to what our general moral duties are. I find such claims unconvincing. They seem, to me, motivated by a certain commonsense aversion against morality being very demanding. As I argue in section 3.4.3, however, a moral theory is not false simply because it is hard to live up to.

1.1.3 A note on methodology

I agree with Singer (2005) that we should aim for convergence on ethical principles informed not simply by automatic intuitions about cases, but by (empirically informed) argument and "reasoned" or "rational intuitions". This does not mean that reasoned argument can settle all (or any) fundamental conflict in ethics forever. At some bedrock level of theorizing, we will always have to defer to affective or non-argumentative components of cognition and decide whether we favor one view over another without being able to justify it, save by appealing to an intuitive response (Weinberg 2016). Having reached bedrock, we might say that some axiom – such as the principle of utility – is "self-evident" (cf. Lazari-Radek and Singer 2014, 149). This, as has been pointed out (cf. Crisp 2014), leaves ethical theory on shaky ground.

Still, it is by way of this intuitionist method that I hope to propagate my moral views; that is, aided by deductive arguments and thought experiments meant to elicit certain intuitions, to try and instill in my peers a feeling that we're onto something.

Part I

Normative framework

2 Rights Talk

[...] contending parties are often fighting for what they think are their just rights; and if we could find a way of arguing which would enable them to agree on what these were, an accommodation would be more easily reached. This could be the philosopher's contribution to world peace, though it would be optimistic to hope that it will be taken up very quickly.

– R. M. Hare (1981, 147)

It is not uncommon for dissidents and disgruntled citizens, targeting some unpopular policy, to invoke *rights* of some kind as justification for their opposition, such as the right to free speech, the right to freedom of movement and freedom from various kinds of coercion, rights to necessary basic goods, to religious, cultural and personal expression, and so on. Hence, exploring what kind of rights claims are being – or could be – legitimately invoked in opposition to climate policies seems a promising aspiration for a thesis on whether moral limits could constrain climate policy. Presumably, those invoking rights to fend off some policy are thereby suggesting where to draw the line, perhaps thinking they have spotted the relevant moral limits.

To assess the legitimacy and relative weight of such rights claims, and to be able to recognize rights claims when we see them, we need an account of what rights might be and how they work. In this chapter, I argue that rights talk is merely shorthand for more basic moral considerations, and that – while the social institution of moral rights serves us well – there are no rights which cannot in principle be set aside to promote a greater moral good. In the terminology of Kagan (1998, 170-175): there are no *deontological rights*. I defend a Consequentialist theory of normative ethics. More specifically: I sketch a rule emphasizing Act Utilitarian position, drawing on Katarzyna de Lazari-Radek and Peter Singer's work on Henry Sidgwick in *The Point of View of the Universe* (2014), R. M. Hare's *Moral Thinking* (1981), as well as Derek Parfit's *On What Matters* (2011a, b, 2017). It is within this normative framework that I will conduct the rest of my explorations in this thesis.

2.1 What are rights?

The language of rights is, as Kagan (1998, 170) notes, "horrendously *ambiguous*" (original emphasis). There's an awful lot we might refer to when applying the term "rights". We might be referring to *legal rights* derived from civil law, *customary rights* grounded in convention, or – as will be my primary concern – *moral rights* grounded in moral reasons (Wenar 2015, 3).[1] Furthermore, we might use "rights" in a very *broad* sense, meaning something akin to "having moral standing as a relevant member of the moral universe", without thereby specifying what that entails in terms of particular, more *narrowly* construed rights (Kagan 1998, 170-171).

Wesley Hohfeld's (1917, 710) seminal taxonomy of the four basic, formal components – or "atomic incidents" (Wenar 2015, 8) – of rights goes something like this:

A *claim* is a first-order relational right which places correlated duty on others.

A *privilege* (or *liberty*) is a first-order non-relational, discretionary freedom, i.e., being free from any duty to the contrary.

A *power* is a second-order right to alter someone's status as a rights-holder (including yourself).[2]

Lastly,

an *immunity* is a second-order right not to have others alter your status as a rights-holder (Wenar 2015, 4-7).

---

[1] Paginations for any *Stanford Encyclopedia of Philosophy* (SEP) reference cited in this thesis will be from the corresponding single column PDF file downloaded via the Friends of the SEP society (https://leibniz.stanford.edu/friends/).

[2] Examples of using a power could be: (1) ordering an employee to do something which it had previously been their privilege not to do; (2) waiving your own claim to privacy by inviting someone into your room; or (3) stripping some under-performing officer whom you outrank of their power-rights, transferring these powers to an eager, promotion-deserving junior officer.

The *privilege* and the *power* can be construed as "active" rights, concerning the rights-holder's own actions, whereas the *claim* and the *immunity* pertain to the actions of others, and are therefore "passive" rights. A further distinction can be made between "negative" and "positive" passive rights, where the holder of a negative right is entitled to not be interfered with, and the holder of a positive right is entitled to some external aid or provision (Wenar 2015, 10-11). Some rights can also be characterized as "general", giving the holder a claim against everybody else, while others are "special", pertaining only to some specific set of agents and/or patients standing in particular relations to each other (Kagan 1998, 172).

Since I am trying to demarcate alleged moral limits on government action, my main concern will be with claim- and immunity-rights, i.e., with what moral duties political authorities might have towards those affected by their actions, especially duties to refrain from implementing some policy which might affect them adversely. However, given the often complex interplay of these Hohfeldian incidents – the claim, the privilege, the power and the immunity – to generate composite rights claims, all of them will be implicated in what follows (Wenar 2015, 7-8).

2.1.1 Are rights absolute?

So much for what *form* rights take. A more pressing concern is the question of where such rights are supposed to get their apparent *force* from. Why does asserting, heeding and enforcing rights claims matter? What are the moral reasons grounding moral rights, to the effect that some rights claims are commonly – and perhaps even legitimately – played as "trump cards", tipping the scales of moral deliberation in favor of the rights-holders in question?

A different way of cashing out the idea of rights as trump cards, is via the notion of moral *constraints*: the idea that some actions are morally forbidden, no matter the consequences (Kagan 1998, 72). Rights, then, correlate with certain moral constraints on conduct: constraints against blocking relevant privileges, against not respecting relevant claims and observing corresponding duties, against exceeding one's powers and against violating the immunities of others. Rights are trumps insofar as these constraints apply even when violating them would serve some greater moral or non-moral good, partial or impartial.

Now, one need not be an *Absolutist* about constraints, and therefore, correspondingly, one need not be an Absolutist about rights. Many reasonable people have adopted the view that rights might serve as trumps only up to certain thresholds (Nagel 1979b, 56; Harel and Sharon 2011, 849-850). If enough is at stake, infringing someone's rights might be morally permissible. Frederick (2014, 376) provides the following paradigmatic example (borrowed from Joel Feinberg):

> HIKER. A hiker on a back-packing trip in the high mountain country is beset by an unanticipated blizzard which strikes the area with such ferocity that her life is imperiled. She stumbles onto an unoccupied cabin, locked and boarded up for the winter, clearly somebody else's private property. She smashes a window, enters, and huddles in a corner for three days until the storm abates. During this period she helps herself to her unknown benefactor's food supply and burns his wooden furniture in the fireplace to keep warm.

The hiker's actions seem justified, even if she's infringing on the property rights of the cabin owner. Clearly, the constraint against breaking into other people's cabins and appropriate their resources does not apply when the alternative is risking your life.

On one construal of this case, the hiker's right to life (or perhaps her right to self-defense) trumps the cabin-owner's right to exclude others from the property. Note, however, that many want to claim that the cabin-owner's right to their property is still intact, even as it is legitimately infringed by the hiker. On what Oberdiek (2008, 127) calls a *general* conception of rights, a right's *content* – what kind of behavior it condemns or condones; what actions it "stands against (or for)" (ibid.) – is not context-sensitive, but an invariable given. Thus, my right to property contains a static composition of Hohfeldian incidents, even as it is overridden by other normative considerations in necessity cases (such as HIKER). In other words: I somehow have the same privileges, claims, powers and immunities under exceptional circumstances as I would normally have, even when relevant others are *not* morally required to align their behavior with the corresponding constraints under these circumstances.

This general conception of rights, construing rights as having context-*insensitive* content, seems mistaken. The content of a right (i.e., what set of Hohfeldian incidents it instantiates) is not self-evident. Rights are not givens. As Oberdiek (2008, 130-131) puts it: "There is always some *reason* to recognize a right as a right […], rights are not normative primitives" (emphasis added). If we stipulate a right with a certain content, we must provide justification

– reasons – for it. Crucially, such justification of rights seems always to have to be, in a sense, *instrumental* – that is, founded on some more basic considerations of what gives us reason to care about the right in question, whether such considerations be Utilitarian, deontological or what have you (Oberdiek 2008, 131-133).[3] Ultimately, these considerations, grounding the content of the right, should – on pain of inconsistency – amount to the very same considerations justifying the constraints entailed by the right; they just *are* the same considerations. That is, the content of a right, being given by more basic moral considerations, defines the correlating constraints on behavior – on what is permissible.

Assuming, then, that the property right of the cabin owner prior to the blizzard was in fact morally justified, and assuming the content of rights to be context-insensitive, as on the general conception, we arrive at the positively odd (bordering incoherent) conclusion that the hiker could be *justified* in ignoring a *justified* – albeit conflicting – rights claim. That is, the hiker is morally permitted to violate a constraint she's not *really* permitted to violate.

We should reject this view. Rather, it seems the property right as stipulated under normal circumstances was at best a tentative, intermediate conclusion about what moral constraints should regulate our interaction with the cabin owner's property, *ceteris paribus* (Oberdiek 2008, 134). In extraordinary circumstances, however, the *final* conclusion about what is morally permissible, all things considered, might look quite different (ibid.), as indeed it did in the HIKER case.

Lest we make ourselves guilty of rights worship, we should reject the general conception of rights in favor of a *specified* conception of rights. On this Specificationist view, the content of a right is sensitive to changing circumstances, seeing as it is in light of what is morally permissible given the circumstances that we specify the content of rights. Here is Oberdiek (2008, 135) again:

> When rights do make an entrance, on [the Specificationist] view, they do so as conclusions about, and not as potential explanations of, the justifiability of certain actions. The explanation for any action's justifiability or lack thereof lies, instead, in the more basic practical reasons that bear on the situation. Specificationism is thus a component of a wider conception of normative thinking in which it is reasons of all varieties that

---

[3] This is not to say that the more basic considerations underpinning the right in question must give instrumental justification for it relative to some even more basic standard, only that the right itself is justified instrumentally relative to these considerations.

fundamentally determine how to act. Rights are not among these reasons, for rights are based upon reasons.

In a sense, then, *all* rights are absolute, insofar as they represent the culmination of sound moral reasoning about what we ought to do under varying circumstances. There is no *infringement on* rights validated by changing external conditions. Rather, where there are exceptions to what would otherwise be rules of conduct, these are built into the rights themselves. The cabin-owner's absolute right to property specifies that we all have a duty not to break into the cabin *unjustifiably*, and there can be no *unjustified* exception (Oberdiek 2008, 128). However, if we adopt a general conception of rights (which I have argued we shouldn't), *no* right is absolute, for we can always discover or imagine cases where moral reasoning leads us to grant that there are plausible or obvious exceptions to allegedly invariable, static rights (Frederick 2014, 377), especially when considering the prospect of large-scale catastrophes, the mitigation of which might demand infringing on somebody's rights.

A corollary of all this, I think, is that one could possibly do moral philosophy and lead an ethical life without rights talk. The basic moral considerations grounding rights could be expressed in other terms (cf. Hare 1981, 153-154; Kagan 1998, 170). However, we may still have sufficient reason to engage in rights talk as part of our moral and political practices. I will return to this point after considering an objection against the view that rights talk is dispensable.

2.1.2 Do we need rights talk?

Judith Jarvis Thomson (cited in Frederick 2014, 383-384) argues that Specificationist theories of rights cannot guide action, because they require that we first figure out what acts are permissible before we ascribe any rights, and therefore cannot explain or justify any particular action by appealing to rights alone. Now, many Specificationists would of course bite this bullet gladly, but Thomson and Frederick (2014, 384-385) claim that rights can be weighty moral considerations on their own, irreducible to other "non-rights moral considerations", and that any reductionist Specificationist theory therefore fails.

To prove this, Frederick appeals to intuitions about two thought experiments which he thinks cannot be adequately explained by appeal to non-rights moral considerations. First, he considers the case

BURGLAR, in which a burglar burgles for pleasure, but always regrets it sincerely and fully compensates the victims, but nevertheless gains more pleasure from the burgling than the victims were pained (pre-compensation), meaning the whole thing resulted in a net increase of welfare (Frederick 2014, 384).

Frederick thinks the burglar is in the wrong. Furthermore, he claims Specificationist accounts are unable to explain *why* in terms of non-rights moral considerations. Then, he considers two further cases that are assumed to be comparably similar in terms of all relevant non-rights moral features:

DAMAGE, in which a malevolent person A destroys some expensive vases belonging to a despised neighbor B, for the sake of damaging B,

and

COMPETITION, in which a malevolent person C lawfully puts some other despised person D out of business by opening a similar shop in the same location, outcompeting D, for the sake of damaging D (Frederick 2014, 384-385).

Even on the aforementioned assumption that these two cases are relevantly similar with regard to non-rights moral considerations (e.g., net costs and benefits are equal), Frederick thinks DAMAGE entails A owing B compensation, seeing as A has violated B's *right* to not have their vases damaged, whereas D has no similar claim on C in COMPETITION, because D has no *right* not to be outcompeted. Allegedly, Specificationists cannot explain this either, seeing as there is no room for rights to do such explanatory work on their own on Specificationist views.

I don't think these cases prove what Frederick takes them to prove. In the BURGLAR case, it rings true that the burglar's actions are wrong, but simply asserting some rights claim as an explanation for this wrongness still begs the question *why?* One plausible answer can be

found if we ask – as moral philosophers of all schools are wont to do – *what if everybody were permitted to do that?* It seems likely that a society without moral and legal restrictions on this kind of burgling-for-pleasure would not be enforcing an optimific set of rules, seeing as a society in which everybody constantly risked having their property burgled (even when compensated) would be expectably less efficient than one in which that risk was smaller, given some plausible assumptions about humans. In the undying words of Thomas Hobbes (1996, 84): "In such condition, there is no place for industry; because the fruit thereof is uncertain."

Similarly, in the DAMAGE case, compensation is owed (or, at least, A's vase-damaging behavior is *unjustified*, and B might therefore be justified in claiming compensation) because the outcome would be a lot less than optimific if everybody went around destroying the property of people they despised, whereas the COMPETITION case arguably fits better with a set of moral rules that hits considerably closer to (even if they are way off) the optimific-mark – a point even ardent critics of contemporary Capitalism might concede.

Thus, explaining our intuitions about these cases seems quite manageable without having to appeal to normatively primitive rights. Now, Frederick might retort that I have missed the gist of his argument, namely that individual moral judgements about specific situations cannot be reduced to non-rights moral considerations, even if a given system of rights derives its legitimacy from such considerations (2014, 385). Here he seems to be making a descriptive point about how most people think intuitively about morality on a case-by-case basis. In our moral practices, rights talk is usually not "merely shorthand for talk about other moral considerations" (ibid.). Rights take on a status as seemingly bedrock. This, however, does not show that rights are either bedrock or indispensable *as a matter of fact*, only that most people *haven't* dispensed with them (yet).[4] As it happens, I shall give reasons why we maybe *shouldn't* dispense with rights talk shortly.

As to Thomson's initial charge that Specificationism cannot guide action, it seems quite clear that rights might serve us well in moral deliberations about what to do, even if they are merely tentative conclusions about what behavior is justified *ceteris paribus*. Tentative

---

[4] Perhaps because most people – and many moral philosophers, at that – conduct most of their moral thinking on what R. M. Hare (1981, 1997) has called the *intuitive* level, as opposed to the level of *critical thinking*. As will become evident, I am quite sympathetic toward Hare's two-level theory of morality.

conclusions may well guide final conclusions. Whatever moral considerations lead us to conclude that some rights hold under normal circumstances, would presumably make valuable starting points for considering extraordinary circumstances (Frederick 2014, 387-388). The considerations grounding, say, a right to property in times of normalcy are not automatically useless or void in times of emergency, even if they are ultimately outweighed by other considerations.[5]

2.2   The case for rights

On the best theory of rights – the Specificationist view outlined above – rights are (perhaps surprisingly) both absolute and dispensable. Whatever force they may have ultimately derives from the more fundamental (non-rights) moral considerations underpinning them. Their content must therefore be specified in a context-sensitive way, to the effect that they cannot really be infringed on.

Now, I doubt that any of this gives us decisive reason to dispense with rights talk, not even with the kind of rights talk that looks suspiciously like deploying the general conception of rights. Even if cabin-owners strictly speaking don't have a general right not to have their furniture burnt, seeing as the furniture might justifiably be burnt for the sake of keeping unlucky hikers alive through unexpected blizzards, it may well be a good thing that most hikers – who, we must keep in mind, are *not* usually imperiled by blizzards – take cabin-owners to have this general right. Similarly, it is probably for the best that most people consider innocents' right to life as absolute in the non-specified general sense, at least most of the time, lest we risk many moral errors being made – or even regressing into an ever more Hobbesian state of affairs. It may be the case, as Harel and Sharon (2011, 852) notes, that "by codifying exceptions one provides overly strong incentives to abuse the codified exception."

---

[5] There are further objections to Specificationist theories, such as the claim that they cannot explain the "moral residue" calling for compensation (or making amends) for justified rights-infringement, for example in emergency cases. Giving a proper reply is beyond the scope of this thesis, but suffice it to say that I follow Oberdiek in thinking any such residue can be comfortably explained by there being relevant moral reasons for *not* infringing on some right, even if these reasons are neither *sufficient* nor *decisive* (Parfit 2011a, 32), therefore not entailing constraints, given the other relevant reasons at play (Oberdiek 2008, 142-144; Frederick 2014, 388-392).

General rights, then, should be part of our mental and cultural repertoire. They should be recognized as *Rules (of thumb)* concerning what behavioral constraints we must respect. They are "Rules (of thumb)" – that is, rules with capitalized Rs and bracketed qualifications – because we should usually downplay their tentative, intermediate status, even as we recognize this fundamental fact about them.[6] A bigger worry than implementing a system of not fully specified rights, is human hubris. Even if it is true, as I think it may be, that innocents might sometimes be justifiably sacrificed for the greater good in extraordinary circumstances, we should be extremely aware of our own epistemic and psychological limitations with regard to specifying and picking out both greater goods and extraordinary circumstances. Placing most emphasis on the usually commonsensical conclusions about morality that general rights represent – though they are intermediate – and imagining rights as something we must in a sense *infringe on* if (and only if) weightier moral considerations call for it, raises the stakes for drawing radical moral conclusions. Hopefully, this makes us more wary of the possible claims and immunities of others, and less prone to being swayed by demagogues with mistaken notions of what would be best, and what policies would be justified to bring it about, given the circumstances.

However, we should not be rule worshippers. General rights are Rules (of thumb), and in clear-cut cases (a slippery notion at best) where a decisively greater good requires their infringement, they can – or *should* – be infringed. Consider the following stock example, reimagined by Lazari-Radek and Singer (2014, 296):

BOMB, in which the *only* way to prevent a nuclear weapon from wiping out a city, killing, maiming and sickening millions of people, is to torture a terrorist's child.

The child has a general right not to be tortured. But any plausible *specified* right not to be tortured must arguably include an exception clause for cases where your being tortured is the

---

[6] I adapt, here, the term "rule of thumb", despite it possibly being – as pointed out by Hare (1981, 38) – "thoroughly misleading". Hare stresses that breaching what could genuinely be described as a *mere* rule of thumb would spark no intuitive compunction or feeling of regret in the agent. Breaching a moral rule, however – such imperatives not usually being conceived of as mere heuristic devices – will for most people involve some level of anguish, however slight – and *should* provoke such feelings, even, on account of their precautionary effects. Hare suggests the term "*prima facie* principles" as a better locution. Nevertheless, seeing as I want to preserve the word "rule", and because I think any parenthesized qualification should be invoked more as an afterthought than as an introductory salute, I will stick to my "Rule (of thumb)" expression throughout, in the name of its beneficial lexical effects (see Cappelen 2018, 122-134). I hope my use of brackets will prevent any harmful such effects. The reader may substitute "*prima facie* principle" for "Rule (of thumb)" at will.

only way to save millions of people from death and excruciating pain. On the general conception, then, the right can be justifiably infringed in cases sufficiently similar to BOMB.

Your right not to be tortured is a Rule (of thumb) that nobody torture you. Dismissively, we could call this a *mere* rule of thumb, but that would be irresponsible. The Rule (of thumb) against torture should not be taken lightly, for if it were, we might open the door *too* wide for people eager to expand the authority of police and military personnel, and the sum total might be an unjustifiable state of affairs where too many are tortured and a legitimate fear of being tortured spreads to the benefit of nobody. Thus, we have weighty non-rights moral considerations underpinning the view that we should speak about the prohibition on torture in near absolute terms, or otherwise ensure that people respect this Rule (of thumb).

A pertinent question is whether the bracketed qualification belongs in a public discussion about morality at all, given human imperfections. Perhaps it would be optimific if everyone were simply taught the Rule against torture, *sans* qualification. But what then if we were ever to be faced with a BOMB-like situation? This necessitates a discussion of what Sidgwick (cited in Lazari-Radek and Singer 2014, 293-294) called *esoteric morality*, the notion that "secrecy might render good an act that would normally be bad" (ibid.), and that knowledge of what might be right to do in special circumstances – such as BOMB – should itself be kept secret (from most people). It is to this topic I now turn.

2.3   Esoteric morality

There is something deeply unappealing about the idea that facts about what is right to do should perhaps be kept secret to most people, not to mention the related idea that secrecy can be a right-making property of actions under certain circumstances (Lazari-Radek and Singer 2010, 37-42). To hold the former view, one must be committed to thinking that not all people can handle or access moral truths, such as those entailed by the latter view; one must agree that "some people know better, or can learn better, than others […]" (ibid., 35). Consequently, the doctrine of esoteric morality might be cast as nothing but the assertion of the corruptibility of one's fellow, less-than-fully reasonable humans. Esoteric morality, then, reeks of paternalistic arrogance and may lower the bar for manipulation of the public by some

smug elite. Bernard Williams famously called it "Government House" Utilitarianism on such grounds (Williams 1995, 166).

Nevertheless, a doctrine of esoteric morality may well be true. In fact, any viable Consequentialist theory has to allow for some degree of esoteric morality, because there are obvious (and less obvious, but not obviously implausible) hypothetical cases where some degree of secrecy about what is right to do would yield the best consequences (Lazari-Radek and Singer 2010, 37-42). For example, as Lazari-Radek and Singer point out: It could be the case that one small, elite group of law enforcement personnel defensibly could be taught how to use excruciating methods of psychological torture to gain reliable information from terrorists, so as to equip them for BOMB-like scenarios, whereas other, less disciplined law enforcement personnel, prone to applying such methods wrongfully, should be strictly prohibited from treating prisoners in such inhumane ways, and therefore be taught a different code of conduct (and be kept in the dark about the workings of the elite unit) (ibid., 38-39).

Consequentialists, of course, cannot disregard consequences. Even Rule Consequentialists, who think the right-making property of a given act is its alignment with some moral code which would yield the best consequences if it were *universally* accepted, and who wisely include the rule "avoid disaster" (but not, say, "maximize the aggregate good") in such a code, must – to remain Consequentialists – agree that esoteric morality is true if secrecy about how a disaster X ought to be prevented may be necessary to prevent a different disaster Y. This is indeed plausible. Drawing on the work of Brad Hooker and Richard Arneson, Lazari-Radek and Singer (2010, 44-49) give the following imaginary example:

GENIUS DESERTER, in which the world consists of a million imbeciles and only one genius. The imbeciles can only internalize very simple moral rules, whereas the genius wields more complex rules competently and reliably. The only way to ensure that the nation of the one genius and the many imbeciles can retain its military defensive posture against a foreign aggressor, is for the imbecilic members of the armed forces to internalize the rule "stand by your post at all times, unless ordered otherwise". The imbeciles cannot internalize a "prevent disaster" proviso, for they are prone to exaggerated risk analysis, and would desert at the slightest hint of danger. However, the genius – serving in a platoon standing their ground while being overrun by the enemy – can grasp the futility of her particular platoon's situation. Realizing that resistance is futile, and no good will come of

16

her remaining at her post, she runs away. It's better that she lives to fight another day, rather than dying for nothing. She encounters some reporters from a national newspaper asking her why she did not follow the "stand by your post" imperative. To avoid weakening the internalization of this ideal rule – seeing as it is the only thing keeping the imbecilic nation's armed forces from mass desertion, which would be disastrous for the integrity of the nation – she cannot publicly admit to having followed the "prevent disaster" proviso, and instead invents some clever lie, all the while praising her steadfast (now dead) fellow soldiers.

In this example, the genius practices esoteric morality to prevent disaster. Insofar as it is true that there is no utility in futility, and dying for nothing is morally disastrous, Consequentialists should condone the genius' escape.[7] And insofar as mass desertion from a just war is indeed a disastrous prospect, Consequentialists should condone her clever lie as well.

What if we're not Consequentialists? What if – as Nagel (1979b, 54) puts it – we're more concerned with what we are *doing*, rather than what will *happen* upon our doing it? To many thinkers of an Absolutist deontological bent, esoteric morality is an abomination. To these people an act of torture, however secret, is right or wrong *as such*. Typically, an act of torture is deemed wrong, being a blatant and significant harm, and a violation of another individual's integrity and autonomy. To Absolutists, this, of course, means that saving the multitudes by torturing the child in the BOMB scenario is wrong, even if nobody ever learns of it (which minimizes harmful social effects).[8] The constraint against torture is just that: absolute. No amount of secrecy can render torture the right thing to do.

The Absolutist position seems mistaken, even in cases where secrecy does not bear on the outcome. While one might perhaps entertain the thought that torturing a child is not decisively justified by the prospect of saving two, three or ten (or some other relatively

---

[7] One might argue Consequentialists should condone the genius' escape even if her dying would not strictly speaking be a disaster, only decidedly negative. Either way, the "stand by your post" rule would not yield the best result in this situation. Adhering to the ideal rule would produce *no* good consequences and would obviously be disastrous for the rule's sole adherent. Her still adhering to it seems to fly in the face of reason (Arneson 2005, 239).

[8] When I refer to "Absolutists" I mean people who adhere to general, unspecified and relatively simple (as opposed to complex) conceptions of particular rights and moral rules. The *Specificationist* Absolutist (cf. section 2.2) might not think torture wrong in every circumstance, seeing as the content of rights and moral rules would be sensitive to context.

modest number of) adults from being killed or maimed (i.e., there is a constraint against torture up to a certain threshold),[9] the BOMB case seems clear-cut. *One* innocent child can justifiably be sacrificed to save *millions* of people – grown-ups and children alike – from untold suffering and death, if this indeed is the only way we can expect to save them. This imaginary example is farfetched, but instructive. One simply cannot ignore the consequences of actions when deciding what one ought morally to do. Hence, since the relative secrecy of an act will at least partly determine its consequences, ruling out esoteric morality requires a substantive normative argument showing that secrecy can *never* yield an outcome good enough to justify an act which would otherwise be wrong. No such argument succeeds (cf. GENIUS DESERTER above).

Of course, most Deontologists do *not* ignore consequences. However, some still believe that secrecy has no place in morality. On what (Lazari-Radek and Singer 2010, 42-43) calls the "Rawls-Gert definition" of morality (after two of its proponents, John Rawls and Bernard Gert), there is a *publicity condition* for the choice of moral principles. Those who partake in the social contract, and whose behavior is to be judged relative to a given moral code, *must* be able to understand the relevant contract or moral code. What behavior morality stands for or against, and to what degree, must be publicly accessible. Morality, in Gert's terminology, is a "public system" (ibid., 42). The Rawls-Gert definition of morality, then, seems to be: "[Morality is] any rational procedure by which we determine what a community of human beings ought to take as a standard of right or wrong for the voluntary actions of its members" (ibid., 42-43). Brad Hooker, the sophisticated Rule Consequentialist, seems also to subscribe to something like this definition: The ideal code of rules is that which produces the best consequences if accepted be almost everyone, and must be public if almost everyone is to be able to accept it (cited in Arneson 2005, 247).

As Lazari-Radek and Singer (2010, 43) aptly point out, the Rawls-Gert definition of morality rules out the important normative question of whether it can sometimes be right for an individual to do something which will only have the best consequences if the act or its permissibility remains secret – an act which is not condoned by the *public* system of morality. Obviously, that is a proper normative question, calling for a substantive normative

---

[9] We might add, with Parfit (2011a, 229), that the precise threshold could be indeterminate for some moral questions.

argument. It cannot be answered simply by defining morality in a way which rules it out. And, as we have seen, there are cases – however hypothetical – where secrecy about what is the right thing to do may yield the significantly better outcome, and may be justified (e.g., to avoid disaster). This is not to say that publicity isn't *usually* the best way to promote the good (it probably is),[10] only that secrecy about how the good is best promoted may in some cases be the best means to that noble end.

2.4   Rule emphasis versus rule worship: Bracketed Utilitarianism introduced

Morality, then, may be esoteric. The question remains, however, if moral rights *should* generally be conceived of, taught and internalized as Rules simpliciter or as Rules (of thumb)? Should knowledge of justified exceptions to commonly accepted moral rules be reserved for some elite group of ethically competent people (if they can be found or trained)?

I favor Rules (of thumb). While I do believe that some people may in a sense be or become *better at* ethics than others – and, consequently, that these people can be more confident in their ability to identify and act on justified exceptions to general moral rules – I think the intuitive pull in cases such as BOMB cannot be neatly detained by way of selective education. Someone who has been taught the Rule against torture may well be drawn intuitively towards saving the millions by whatever means necessary, even torture, when compelled to consider that grueling choice. Eradicating this intuitive pull by making sure everyone properly internalizes a rule to the contrary would likely come at too high a cost, considering the methods of indoctrination required. The truth will out!

Even if secrecy about the true morality could succeed, I don't think it should. There is no surefire way to guarantee that only a select group of people will be the ones faced with

---

[10] There are many examples of how a Consequentialist rationale could support publicity. Lazari-Radek and Singer (2010, 51-54) list a few: (1) societies require a shared moral code to function well, (2) open discussion about the moral rules implemented serve as a bulwark against some manipulative elite implementing mistaken or self-serving rules, (3) elitism among humans is dangerous, (4) moral education and critical thinking – the lack of which might foster fanaticism (Hare 1981, 174) – would suffer if the true morality was wholly esoteric, (5) it wouldn't be practicable, and (6) the level of trust in society would take a hit if people started to realize that commonly accepted moral rules didn't *really* count.

extreme scenarios like BOMB, should they ever arise. Emergencies do not strike neatly. Other, less dangerous situations not adequately covered by some general moral rule – or involving conflicting moral rules – might also expectably arise, with no elite moral agents present, to the effect that wrong choices are made in the name of inadequate rules (or an inadequate grasp of their foundations). Given that such minor hiccups might in fact occur quite often, it seems odd to presume that it would be better on the whole if most people were kept unaware of the possibility of justified exceptions to moral rules. It seems likely – granting people the benefit of the doubt – that most would be able to honor moral rules under normal circumstances, even when aware of the possibility of the rules changing under exceptional circumstances. The more people ethically equipped to make the best choices under varying circumstances, the better. This rules out strict (context-insensitive) rule worship as a strategy for moral deliberation and decision-making.

It would be disastrous, however, if making the best choice – the choice with the best consequences – were seen as the sole or overriding moral imperative in any and every situation. At least if "best consequences" is cashed out in terms of the most (in)famous, most parsimonious Consequentialist theory, namely *Act Utilitarianism*. The principle of utility as deployed by Act Utilitarians (typically coined as some version of the imperative "maximize happiness") is not the best moral decision procedure, not even relative to Act Utilitarianisms own tenets; it would not maximize happiness if every actual human being were always striving to act in accordance with that rule alone. Interestingly, Katarzyna de Lazari-Radek and Peter Singer – two of Act Utilitarianism's staunchest defenders – agree (Lazari-Radek and Singer 2010, 55). They expound:

> […] 'Maximize the good' is not the best decision procedure, at least not always and not for every person, since it may bring about bad consequences such as a lack of trust, or something even worse. This may be a problem of human nature and its tendency to think of oneself rather than of others – once people are allowed to break widely accepted moral rules in order to maximize the general good, it is entirely possible that they will also break them to achieve their own ends. There is also a problem with calculating what the result will be. Some people are able to calculate well, while some, whether because of a lack of intelligence, or of time, or for other reasons, are likely to miscalculate what would be best in a given situation.

Two points should be added:

Firstly, the epistemic argument against Consequentialist moral decision procedures should be

more forcefully stated than by Lazari-Radek and Singer. Calculating the causal ramifications and ultimate results – the *actual* results, that is, as opposed to merely *foreseeable* consequences – of many (if not most) individual actions is, as it stands, humanly impossible. Virtually nobody is able to "calculate well" when considering a wide range of actions, particularly actions (even tiny ones) which affect the identities of future people, whether directly or indirectly (Lenman 2000). This is not to say that Consequentialism cannot reasonably recommend certain patterns of action, social norms or rules on the basis of their probable systematic effects (Burch-Brown 2014), only that a given individual poised to perform some act cannot hope to arrive at sound conclusions about the ultimate consequences of doing so just then. (I have more to say about this in section 3.5.3.)

Secondly, if the principle of utility were recognized by convention as the one moral rule everyone should follow (all else being equal), and all rights talk were abandoned, powerful people would surely be able to fit many subtle and explicit horrors within the scope of terms like "good" or "useful". Majorities could be convinced that atrocities suffered by minorities were justified by concern for the greater good – however implausible such promises may be. Indeed, twisted, future-biased Utilitarian arguments have historically been deployed to that end, by all manner of tyrants and aggressors (Maynard 2014, 832).

While Act Utilitarianism may well supply the best criterion of right action (we ought to do what maximizes net happiness or utility), the universal adoption of its principles as a guide to action would likely not be best. People should *not* be encouraged to maximize the good all the time, even if this is what they *ought* to be doing. Consequently, it may be true that Act Utilitarianism itself should be kept comparatively secret, if maximum net happiness is what we're aiming for (Jamieson 2007, 168; Lazari-Radek and Singer 2010, 55-56). Every fallible human being needs a more elaborate moral code than the lone principle of utility. However, the moral code must also be digestible; its rules should be internalizable and practicable in a wide range of situations where doing rigorous moral philosophy would be inexpedient (for want of time, competence, will, etc.). This concern limits the level of specification and contextualization possible for a given rule. We are not, in Hare's terminology, omniscient and omnipotent *archangels* (Hare 1997, 138). Therefore, rather simple, general rules – probably quite similar to those familiar from popular morality – would likely serve us best on most occasions (Kagan 1998, 229-230).

Yet, for reasons already given, we need a moral code which allows for its being breached in special cases – most obviously in certain disastrous emergency scenarios, but plausibly also in cases where sticking to the general rule would do absolutely no good. Whether adherence to an ideal rule is actually for the better in a particular situation will often depend on the level of compliance then and there (Arneson 2005, 237). In some cases of widespread non-compliance, the individual might achieve nothing valuable (but sacrifice much of value) by heroically complying nonetheless (ibid., 239).[11] A moral theory requiring pointless sacrifice in the name of some ideal code seems manifestly *un*-ideal, if not flat out mistaken.

Thus, we need the moral flexibility provided by Act Utilitarianism for special situations, even as we stick with some expectably optimific list of general, rather simple moral rules in our day-to-day lives and most non-emergency situations – perhaps even most emergencies (a point I shall return to in section 5.2). This is what leads to the recommendation of formulating morality as a set of Rules (of thumb). We should not be rule worshipping Utilitarians. We should, however, be *rule emphasizing* Utilitarians. In fact, we should be Rule emphasizing (Utilitarians). Rather than keeping Utilitarianism a secret, it should be kept in brackets; always present as a last resort *in extremis* – e.g., when Rules (of thumb) conflict, or in novel situations where no ready-made Rule (of thumb) presents itself – yet almost never dominating our moral deliberations.

The *Bracketed Utilitarianism* I am proposing may serve to remind us that no Rule is ultimately above its consequences, but also that no individual decision-maker is entitled to just bypass the Rules and reach directly for Utilitarianism through the brackets. Even competent moral agents – geniuses, at that – should keep Utilitarianism in brackets, because *everyone,* except archangels, is an imbecile with regard to foreseeing remote consequences. (Or, if that's too harsh, at least no one is epistemically justified in considering themselves significantly better than anyone else at getting Utilitarian calculations concerning the ultimate effects of their actions *exactly* right.) Also, the geniuses' example may itself be decidedly negative given the imperfect nature of likely onlookers. When confronted with novel situations that seem to require the deployment of Act Utilitarian principles, even geniuses should – circumstances permitting – take extra care upon deliberating and follow some

---

[11] This goes even for second-order rules specifying what one should do given the actual level of non-compliance to the first-order ideal code, so long as these second-order rules are also formulated as answers to questions of what rules would yield the best consequences if everybody accepted and followed them (Arneson 2005, 241).

general Rule (of thumb) if applicable. If not, moral mathematics it is – but not with brash self-confidence or arrogance. While Utilitarianism provides the best criterion of the right, it is not typically the best (and therefore not the right) decision procedure (Sinnott-Armstrong 2015, 16-17). Correctly identifying the exceptions is no small feat. The Utilitarianism we most need in practical deliberation is a humble and careful one, emphasizing rules and bracketing itself, but ready to get the job done when push comes to shove.

2.5   The moral force of rights talk

So far, pertaining to the question of where rights get their moral force from, I have said the following:

(1) Rights are not normative primitives. They are founded on more basic moral considerations.

(2) At the foundational level, Act Utilitarianism may be the theory which yields the most plausible basic moral considerations and specifies the correct criterion of the right.

But...

(3) Act Utilitarianism is not viable as a decision procedure or guide to action in most cases, and should be bracketed, given the way humans and the world currently are.

(4) Our moral decision-making should largely be guided by some expectably optimific list of Rules (of thumb), Act Utilitarian calculation being reserved for special cases – preferably as a last resort.

Moral rights enter the picture around (4), amounting as they do to certain Rules (of thumb) constraining behavior. I have not yet thoroughly defended (2) and will do so in the next chapter.

The arguments in favor of (2) which I have hinted at so far amount to:

(5) We need a moral theory which can explain reasonable (and, I think, common) intuitions about emergency cases such as BOMB, where a (significantly) greater good calls for breaking a commonly accepted moral rule.

(6) We cannot ignore consequences when deciding what we ought morally to do here and now; we should not be rule worshippers but take into account the particular circumstances of our actions.

Together, these considerations rule out both deontological Absolutism (Kagan 1998, 79) and iterative Rule Consequentialism (Arneson 2005, 242). But there are several other contenders Act Utilitarianism still has to face. In the next chapter, I will present the positive case for Act Utilitarianism as well as give a brief defense against some of the more urgent objections pressed against it.

3 Utilitarianism and the Levels of Moral Thinking

Ethics at large may be defined, the art of directing […] actions to the production of the greatest possible quantity of happiness, on the part of those whose interest is in view.

– Jeremy Bentham (2010, loc. 4774)[12]

Define Act Utilitarianism as the theory that some act is morally right if and only if this act – among all available acts – is the one which maximizes net utility, i.e., the act which yields the best possible consequences, taking into consideration all relevant members of the moral universe – everyone, that is, human or non-human, who are (or are capable of being) affected by the act, however indirectly. Act Utilitarianism, then, is a Consequentialist theory. Consequentialist theories have to be combined with some plausible account of which consequences are best to produce substantive moral arguments; we need a theory of the good. In what follows, when using the term "Utilitarianism", I will be referring to Consequentialism coupled with monism about the good, specifically Hedonism – the view that pleasure (widely construed, in all its manifestations) is the only intrinsic good, its absence or opposites being correspondingly bad (Sinnott-Armstrong 2015, 3). In this chapter, following some preliminary comments on the different levels of moral thinking, I shall be defending Act Utilitarianism – thus defined – against some of its more prominent opponents.

3.1   Moral factors and moral foundations

Following Kagan (1998, 17-21), we might assert that normative ethics may be done on two interdependent yet distinct levels: the *factoral* level and the *foundational* level. On the

---

[12] Paginations for Kindle e-book editions are given as location numbers (prefixed "loc.") when original page numbers are not given in the e-book.

factoral level, we are trying to answer questions about which factors bearing on a situation are morally relevant, and how these factors interact to yield verdicts about what we ought morally to do. On the foundational level, we're dealing with questions about what explains the moral significance of the relevant factors.

Subscribers to Consequentialism on the factoral level argue that the only morally relevant factor to consider when pondering some course of action is its consequences. Other candidate factors – such as duties, special obligations, intentions, desert, demandingness, options, rights, etc. – are only relevant insofar as they contribute to the net goodness of the outcome, and have no intrinsic moral relevance (Kagan 1998, 60). Consequentialism on the foundational level explains the relevance of whatever factors are in fact relevant by appeal to the intrinsic significance of maximizing (or perhaps satisficing) the good overall (as opposed to what is good for only *some*, the evaluative focal point of the Egoist) (ibid., 213).

Kagan makes much of the fact that Consequentialism on the foundational level might not yield consequentialism on the factoral level. *Satisficing Foundational Consequentialism*, for instance – where one is only morally required to increase net utility up to a certain threshold – might yield a normative theory on the factoral level which grants significance to *options*, leaving room for supererogatory acts of self-sacrifice which are not morally obligatory (ibid., 220). He stresses, however, that *Maximizing Act Consequentialism* on the foundational level must yield Consequentialism on the factoral level as well, because admitting that whatever moral relevance some factor might have derives from its contribution to the goodness of the outcome, amounts to reducing deliberation on the factoral level to questions of maximizing good outcomes (ibid., 115-223).

This, however, does *not* entail that asserting Maximizing Act Consequentialism at the foundational level, as I am wont to do, requires always (or ever) promoting the kind of moral deliberation which Consequentialism at the factoral level invites. To see this, we must direct our attention to another separation of moral thinking into two levels, succinctly articulated by R. M. Hare (1981, 1997): the *intuitive level* versus the level of *critical thinking*.

26

3.2 Two-level Utilitarianism: Intuitive versus critical thinking

The intuitive level of moral thought should be familiar to most. It is where we – in our daily moral practices – draw on relatively simple principles instilled in us by culture and society, through upbringing and schooling (or by ourselves, through some self-initiated process of moral improvement), to decide what we ought morally to do. This process is quite effortless and automatic, until some such time when intuitive principles come into conflict, or none are applicable. At this point, rational moral discourse tends to falter or break down.

When such conflicts between intuitive moral principles arise, critical thinking should – but often doesn't – ensue. At the level of critical thinking, one must – by appeal to reason (Hare 1981, 32) or reasoned intuitions (see Singer 2005) – try to reach a final conclusion about what ought to be done, all things considered.[13] Recall the disagreement between adherents of *general* versus *Specificationist* conceptions of rights (chapter 2): General rights belong on the intuitive level, where they may collide often – two conflicting principles being apparently justified at once. Specified rights, on the other hand, belong to the domain of critical thinking, where, ideally, justification is never partial, *pro tanto* or *prima facie*. Hare (1981, 26) gives an elucidating example:

> The critical level is that at which the minister was operating who put a placard on the 'wayside pulpit' outside his church in Yorkshire [...] saying 'If you have conflicting duties, one of them isn't your duty'.

Critical thinking aspires to reach final conclusions. On Hare's account, this process "consists in making a choice under the constraints imposed by the logical properties of the moral concepts and by the non-moral facts, and by nothing else" (Hare 1981, 40). We need not agree with this, even if we agree that there is a level of moral thinking where apparent moral conflict can be solved, and where moral principles are selected on their merits by way of reason – or at least some cognitive procedure more rigorous than mere prejudiced intuition.

---

[13] Hare has a rather strict view on intuitions, claiming that no moral intuitions of substance can be appealed to in critical thinking (Hare 1981, 40). Singer (2005) is less categorical. He restricts himself to banishing intuitions which may be explained away by pointing to lurking variables like upbringing or evolution.

Kagan's Factoral Consequentialism is something one arrives at by way of critical thinking, for example by bringing oneself, through reasoned deliberation, to the conclusion that Maximizing Act Consequentialism might be true at the foundational level. The question of whether Factoral Consequentialism demands the adoption of Consequentialist principles at the intuitive level of moral thinking (by everyone at all times), however, is a separate matter. Will this likely produce the morally optimific set of acts? I argued to the contrary in section 2.4.

I elaborate on this point to underscore that the Act Utilitarianism of which I am in favor does not necessarily prescribe widespread Act Utilitarian calculation. It is a sort of *two-level* Utilitarianism, reconcilable with retaining much (perhaps even most) of our current moral thinking at the intuitive level, while deploying Act Utilitarianism as the tie breaker and selector of principles – or Rules (of thumb), as I have called them – at the critical level.

I will now give an exposition of the arguments which have convinced me that Act Utilitarianism – akin to the Classic Utilitarianism of Bentham, Mill and Sidgwick – is the most plausible theory of normative ethics at the foundational level. Such theories are made up of several distinct claims (Sinnott-Armstrong 2015, 2-4), and I will deal with the most important ones separately. First, I'll defend Consequentialism in general. Then I'll argue that the best version of Consequentialism seeks to maximize the amount of good actual consequences (e.g., from acts) directly, and I'll defend Hedonism as the best candidate theory of the good. Lastly, I will try to answer some important objections.

3.3   The case for Consequentialism

Can morality be reduced to questions about consequences? When assessing what we ought morally to do, or the normative merits of some already performed act, should we give no weight to the characteristics of the act itself, or the intentions and motives behind it, but only consider its resultant effects? Consider:

> OBSCENE MURDER, in which a wicked murderer gruesomely kills an innocent woman
> because it gives him obscene pleasure, not knowing that the woman would otherwise, had

she not been killed, have gone on to birthe and raise some even more murderous tyrant – the next Hitler, the ravages of which the world is spared because of the obscene murder.

Even if the world is relieved of the presumably harrowing strain of having to suffer another Hitler, and the OBSCENE MURDER therefore arguably nets a positive result, that was not why it was committed. The wicked murderer's act was entirely self-serving and cruel. Nor is the good news ever brought to anyone's attention. No one is in a position to know that the woman's murder has this positive result. The affair looks utterly tragic to any actual bystander, such people being unable to fully assess the consequences – this being the privilege of archangels (or philosophers inventing imaginary examples from swivel chairs). Yet the actual consequences are what they are, and the obscene murder, according to Consequentialism, was permissible. This seems a troubling conclusion. To dampen the blow, however, we can help ourselves to some Parfitian distinctions.

3.3.1 Senses of moral wrongness

Derek Parfit (2011a, 150) distinguishes four different senses of moral wrongness pertaining to situations where some agent must act without knowing all the morally relevant facts: *fact-relative*, *belief-relative, moral-belief-relative* and *evidence-relative* wrongness. These are all defined in terms of what he calls the "ordinary" sense of moral wrongness. Parfit thinks an act is aptly labeled (morally) "wrong" in the ordinary sense when it is disallowed by some principle of morality and is performed by someone who knows all the morally relevant facts.[14] The four senses of "morally wrong" Parfit urges us to adopt *in addition* to the ordinary sense are all defined in terms of the ordinary sense and pertain to agents who do *not* know all the relevant facts upon acting (i.e., most of us, most of the time). According to Parfit, an act is wrong in

(1) the *fact-relative* sense, when the act would be wrong in the ordinary sense if we knew all the morally relevant facts,

(2) the *belief-relative* sense, when the act would be wrong in the ordinary sense if our beliefs about the morally relevant facts were true,

---

[14] Parfit's *Triple Theory* – his take on the fundamental principle of morality – states that an act is *in fact* wrong "just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable" (Parfit 2011a, 413).

(3) the *evidence-relative* sense, when the act would be wrong in the ordinary sense if we believed what the available evidence gives us decisive reasons to believe, and these beliefs were true (ibid.),

and

(4) the *moral-belief-relative* sense, when the agent *believes* the act to be wrong in the ordinary sense (ibid., 157).

If Consequentialism is true, the OBSCENE MURDER is the right thing to do in the fact-relative sense, seeing as the only morally relevant facts are facts about resultant effects and the resultant effects are ultimately positive, all things considers, because there is no new Hitler. However, the murder remains wrong in the belief- and evidence-relative senses, and plausibly in the moral-belief-relative sense as well, depending on the wicked murderer's moral beliefs. It is wrong in the belief-relative sense because people believe that the murder results in tragedy overall, and if that were true it would be wrong in the ordinary sense. It is wrong in the evidence-relative sense because the visible consequences of the murder arguably provide decisive reason to believe that the act is wrong,[15] and if that were true it would be wrong in the ordinary sense. Finally, it might be wrong in the moral-belief-relative sense if the wicked murderer was at least half-decently raised and had adopted commonsense moral views, in spite of which the murder is knowingly committed.

Thus, distinguishing the above senses of moral wrongness, we remain entitled to deem the OBSCENE MURDER wrong in some relevant senses, even if Consequentialism bars us from judging it wrong in the ordinary and the fact-relative sense.[16] The question remains, however, whether or not facts about consequences are the *only* morally relevant facts capable of rendering acts wrong in the fact-relative sense. Hare (1997, 163) provides some tentative reasons why they might be. He even proposes that a proper formulation of Consequentialism makes it "hard to see how anyone, Kant included, could fail to be a consequentialist".

---

[15] But see Lenman (2000) about the relative significance of visible consequences.
[16] Judging any actual act, performed in the real world, wrong in Parfit's ordinary sense (i.e., in light of *all* morally relevant facts) is, I think, a task of gargantuan epistemological difficulty for any Consequentialist (see section 3.5.3).

3.3.2 Hare's argument for Consequentialism

Hare's (1997, 164) argument goes a little something like this:

First premise: Acts are morally right or wrong insofar as their morally relevant defining properties make them so.

Second premise: An act is defined by how it influences the course of events causally, i.e., by its consequences (e.g., to kill is to *cause* death). In other words, the defining properties of an act are its consequences.

Conclusion: Acts are morally right or wrong insofar as their morally relevant consequences make them so.

One might object to the second premise by maintaining that acts are also defined by the intention behind them. Hare (ibid., 165) parries such blows by showing how even Kant's candidate for what is good without qualification, namely a good will (i.e., good intentions), is only good insofar as it wills good consequences:

> When we are wondering what intention to form, the intentions that are the possible candidates are all intentions to bring about certain consequences; that is, to do certain actions or to make the course of events different in certain ways. So the will itself, which is being formed in this deliberative process, is a will to bring about certain consequences. They are what is willed—the objects of volition, as Kant calls them. So, although the only good thing without qualification is a good will, what makes it a good will is what is willed [...], and that is the consequences that are intended.

A will willing vile acts of villainy would not be a good will. Good wills must necessarily will (actually) good consequences, otherwise they won't be good in the fact-relative sense. That is, mine would not be a good will if I willed the violent death of some innocent baby, even if I was of the delusional belief that this event would please some god or other, or otherwise net a positive result. The poor baby's demise would, as a matter of *fact*, be a bad consequence to will – its life being wasted to the detriment of all who knew and never knew it (assuming, of course, this particular baby would not itself be – or in some way bring about – the next Hitler, or in any other way spur some perverse chain of events by not being killed, and I harbored no such belief). My will in this imaginary example might well be good in some other sense, such as the belief-relative sense, but not, I submit, in the (weightier) ordinary and fact-relative

senses. Furthermore, this example generalizes to less extreme cases: a will not willing the best possible result is not the best possible will.

So far, nothing is settled as to *what* consequences are morally good, bad or even relevant to discussions of morality at all. Consequentialism needs a theory of moral value – of what the good *is* – to be of much use (Railton 1984, 148). Then there are questions of agent culpability with regard to distant effects and the relation between actual and expected effects, which are a matter of some contention (Sinnott-Armstrong 2015), and to which I will return. Still, Hare's tentative Consequentialist conclusion is instructive. It can be summarized in the following question: If some act or thought has no consequences whatsoever – actual or intended – how could it possibly be morally right or wrong, or even be an act or a thought at all? Morality, it seems, hinges entirely on consequences.

3.3.3 The paradox of doing less good than we might

In proposing the above argument, Hare is "consequentializing" a self-declared non-Consequentialist school of moral thought (Brown 2011); showing, that is, that Kantians are actually, unbeknownst to themselves, closeted Consequentialists. Some theorists assert that *every* plausible moral theory is susceptible to such consequentialization (Portmore 2007, 39). Even if that might be an overstatement, Dreier (1993, 24) highlights the appeal of such a consequentialization strategy. There is, he observes, an "ugly stigma" attached to the rejection of Consequentialism, from which one can escape by admitting that "every moral view is consequentialist". Dreier (ibid.) portrays the "process of stigmatization" like this:

> The consequentialist asks whether we don't, in fact, take the happiness of others to be a good thing. Of course, we do, how could one not admit that it is a good thing that others are happy? Then, the consequentialist asks, are there circumstances wherein one ought not to promote the (all considered) good, under which we ought to prefer less good to more? Common sense morality seems to say that there are. But how could this be? The rules that constrain us from doing more good need strong justification to overcome the paradoxical air of requiring us to do less good than we might.

Trying to explain the importance of rules in terms of their good consequences will do the commonsense moralist little good; "[...] then the consequentialist has us in his net" (Dreier 1993, 24). Neither is simply asserting the intrinsic importance of following rules an attractive argument. How could following rules be important, if not because doing so has some positive effect on the chain of events (cf. Hare's argument)?

Yet, commonsense moralists typically want to reject having to justify rules in light of their resultant effects, seeing as such justification allows for breaking rules to prevent even more severe rule breaking. "We are not willing to twist a child's arm even if the information we could thereby get from his grandmother might help us to save two other children from arm-twisting" (ibid.). But this leaves the commonsense moralist in a "very uncomfortable position" indeed, having to resort to seemingly dogmatic rule worship.

> On the one side of deliberations is the good; we ourselves admit that it is good, and feel thus committed to favoring its advancement. On the other side are mere rules. Our rejection of consequentialism leaves an ugly stigma (ibid.).

The commonsense moralist can, Dreier (1993, 25) thinks, escape the "paradoxical air" of such views by admitting that she is, in fact, out to maximize the good, yet denying an *agent-neutral* understanding of the good in favor of an *agent-centered* or *agent-relative* conception. A theory is agent-relative insofar as it provides different agents with different aims; otherwise it is agent-neutral (Brown 2011, 761). Thus, a commonsense moral theory with an agent-relative conception of the good – let's take as our example some version of Virtue Ethics – would, if consequentialized, prescribe some rule such as this: Always strive to be courageous when courage is required, because failing to be courageous when apt is a negative consequence. Two separate agents, A and B, applying this rule – both aiming to be courageous – will nevertheless have *distinct* aims. A's aim will be to muster *A's* courage, B's aim to muster *B's* courage. In some situations, A's being courageous might preclude B's being courageous (Dreier 1993, 37). While agent-neutral theories (e.g., Act Utilitarianism) would spur everyone to bring about the same overall result, aligning the theoretical goals applying to every agent (e.g., maximizing total happiness), agent-relative theories (e.g., some version of Virtue Ethics) might prescribe an agent with an aim which conflicts with another's theory-given end (ibid., 22), even as both kinds of theory aim at maximizing the good (e.g., the promotion of individual virtue.)

Consequentializing your moral theory, then, seems a promising escape route is you want to elude the paradox of being required to do less good than possible. In promoting the good to the best of one's ability – even if the relevant conception of the good is agent-relative – one

could not have done any better, so to speak, and the apparent paradox dissolves. All the better for Consequentialism.[17]

3.3.4 The explanatory force of Consequentialism

The above arguments may seem fishy to some. Consequentialism might look less like a substantive alternative theory, and more like a set of vacuous claims, if it simply tries to assimilate almost every competing account of morality (Brown 2011, 750). To win over stragglers, I'll conclude this section with what is perhaps the most *intuitively* forceful argument for Consequentialism: Consequentialism alone, it seems, can provide *non-arbitrary* justification for thresholds between right and wrong in some important cases, as well as non-arbitrary criteria for ranking alternative options for action when *pro tanto* moral rules conflict. Thus, Consequentialism better explains some common moral intuitions than do its deontological counterparts (Harel and Sharon 2011, 851; Sinnott-Armstrong 2015, 32-33).

Consider again the BOMB case. A Deontologist might want to say that torturing someone to save the lives of, say, *five* people is wrong. But intuition tells many of us (even some Deontologists) that torture, however gruesome an act, is *not* automatically wrong in BOMB-like scenarios, where *millions* of people could be saved if some individual – a child, even – were subjected to it. Consequentialism explains why: In the BOMB case, torturing the individual – out of all available options – yields the best aggregated consequences. A Deontologist, however, will be hard pressed trying to justify some threshold at which torture suddenly goes from being forbidden to being permissible or mandatory, especially if reluctant to appeal to the relative goodness of outcomes.

Similarly, our Deontologist, accepting some general rule like "never break a promise", might find herself in a situation where two made promises come to preclude each other. It seems obvious to her that breaking one of the promises will have worse consequences than breaking the other, but – being a *hardcore* Deontologist – she can't in good conscience defer to the relative goodness of outcomes, though it is an obvious tiebreaker (Sinnott-Armstrong 2015, 33). Had she been a Consequentialist, she'd not only intuit which promise to break, but she'd

---

[17] But see Brown (2011) for an argument that agent-relative theories – and some other theories – may resist consequentialization.

be able to argue the case with some authority. Consequentialism, then, packs quite an explanatory punch.

3.4   What kind of Consequentialism?

Assuming that some version of Consequentialism is true, what version is it? Sinnott-Armstrong (2015, 3-4) identifies ten logically independent claims on which Consequentialists might take differing stances, which can combine in different ways to make different versions of Consequentialism. Classic Act Utilitarianism accepted them all. The claims are:

1. *Actual Consequentialism:* Only *actual* consequences decide whether an act is morally right (i.e., intended, likely, foreseen or foreseeable consequences do not count).

2. *Direct Consequentialism:* Only the consequences of *the act itself* count, as opposed to consequences of generally adhering to some rule covering such acts, the consequences of the agent's motive, and so on.

3. *Evaluative Consequentialism:* Only the *value* of the consequences count, as opposed to non-evaluative features of consequences (e.g., features not involving or affecting subjective judgement or experience).

4. *Hedonism:* The value of the consequences is determined solely by the net sum of *pleasures* and *pains* in the consequences. Other supposed goods (freedom, knowledge, honor...) do not count.

5. *Maximizing Consequentialism:* Some act is what we ought morally to do if and only if its consequences are the *best possible* consequences, as opposed to it being merely satisfactory or incrementally better than some other option.

6. *Aggregative Consequentialism:* Which consequences are the best is determined by summing or otherwise aggregating the *parts* of the consequences – i.e., the value of a world is determined by its constituent *local* phenomena (e.g., experiences, lives or

societies) (Bostrom 2011).

7. *Total Consequentialism:* Only the *total* net good in the consequences counts toward moral rightness, as opposed to the average good per individual (Sinnott-Armstrong 2015, 3-4).

8. *Universal Consequentialism:* The consequences for *all* sentient (or otherwise morally relevant) beings count toward moral rightness.

9. *Equal Consideration:* All those individual beings who count as morally relevant count equally (the queen's pleasure is as valuable in isolation as the pauper's).

10. *Agent-neutrality:* Whether consequences are evaluated from the perspective of the agent or an observer does not determine their value.

I will defend a theory similar to Classic Utilitarianism, in that it accepts most of these claims. Reviewing all ten is beyond the scope of this thesis, however, so I will limit myself to defending some of the more important and controversial claims: *Actual Consequentialism* (1), *Direct Consequentialism* (2), *Maximizing Consequentialism* (5) and, at last, *Hedonism* (4). In the final sections of this chapter, where I reply to some objections, I will be touching upon some of the other claims as well, e.g., *Total Consequentialism* (7).

3.4.1 Actual Consequentialism

Does Consequentialism require that we maximize *actual* or *expected* good consequences? If the former is true, one could be acting with the best of intentions – even feeling quite confident that the act would yield a positive result – and nevertheless end up acting wrongly, perhaps because the act unexpectedly causes some freak accident. However, condemning such well-meaning acts as immoral, *blaming* the agents performing them, seems unduly harsh. Unfair, even. Thus, some have concluded that agents could only be obliged to maximize the good consequences they can reasonably *expect* to bring about (Jackson 1991). Similarly, some argue that agent responsibility wanes with time and distance, like "ripples on a pond" (Smart 1973, 33), to the effect that we cannot really be held responsible for actual

future ramifications of our choices, insofar as the relevant events are distant and not (potentially) visible to us.

Persson (2008, 351) distinguishes between an "internal" and an "external" side of any intentional action, the *internal* side being what we decide to do or try to bring about, the *external* side being the consequences we actually end up bringing about – whether or not we intended to. The external side of actions concerns *actual* consequences, while the internal side concerns what we expect or intend. This is an important distinction, because it allows us to keep questions of *blame* apart from questions of what is morally right or wrong in the fact-relative sense. As Lazari-Radek and Singer (2017, 77) note, the question of whether someone ought to be blamed for their actions is itself a normative question, where the answer – assuming Consequentialism – depends on whether expressing blame ultimately yields the best result. It may well be that we ought to praise someone acting from benign motives with good intentions, even if their particular actions on the occasion have negative ramifications, because such motives and intentions, if cultivated and cherished throughout society, may tend to yield more good overall than alternative motives or intentions.

Yet, as Persson (2008) convincingly argues, there is a sense in which, even if we might not want to blame someone for trying to maximize only expected good consequences – e.g., when faced with situations where there is great risk associated with trying to maximize *actual* good consequences, while the probability of success seems a lot greater if the agent tries to bring about some less than optimal alternative – we would still *wish* for the optimal result. Hence, if someone acted to maximize *expected* utility, but got lucky and ended up maximizing *actual* utility, we'd have to admit that this happiest of outcomes was indeed what *ought* to have happened, externally speaking. It was impartially and objectively – "from the point of view of the universe", as Sidgwick put it (cited in Lazari-Radek and Singer 2014, 133) – the best possible outcome, towards which Consequentialism compels us.

Blame, then, pertains to the *internal* (or subjective) side of actions, and is something we may or may not prescribe depending on whether the relevant internal features of the actions in question are blame- or praiseworthy (i.e., whether blaming or praising these features will tend to have good consequences). The external side of actions (i.e., their actual consequences, whether foreseeable or remote), on the other hand, are regulated by an objective *criterion of rightness* (Persson 2008, 351). This criterion determines the moral value of consequences and

says which of them ought to be brought about, but does *not*, however, determine a *decision procedure*. Assuming a Consequentialist criterion of rightness, any fallible agent – lacking the omniscience of archangels – might be deemed morally irresponsible if internally aiming at what is externally best in some situation where success seems highly improbable, while some decidedly negative result seems the likely outcome. Yet, if by chance the gambit succeeds, and the optimal outcome is achieved – despite what could be reasonably expected – this is indeed what ought to have happened.

Expressed in Parfitian terminology: In the fact-relative sense, we ought to maximize *actual* good consequences. In the belief- or evidence relative senses, however, we ought to do – or try to do – that which is *expectably* best. Assuming, then, that the fact-relative sense of ought is indeed *weightier* than the belief- and evidence-relative senses – in that its content is deduced from the abovementioned objective criterion of rightness (cf. Persson 2008, 353) – we arrive at the conclusion that Actual Consequentialism is true; we ought to maximize actual good consequences, even if we sometimes could be blamed for trying to do so.[18]

3.4.2 Direct Consequentialism

Direct Consequentialism is, on Sinnott-Armstrong's (2015, 3-4) definition, the thesis that only the consequences of *the act itself* determine its moral value. As Louise (2006, 66-67) notes, however, there is some confusion as to the definition of "Direct Consequentialism". Some seemingly take it to mean Act Consequentialism. i.e., the direct application of the criterion of rightness in the evaluation of any set of available *acts*. There are, however, many other possible "evaluative focal points" (Kagan 1998, 214), such as rules, motives, dispositions, beliefs or character traits. Thus, a more accurate definition of Direct Consequentialism would be: "[...] those Consequentialist theories that apply the criterion of rightness directly in the evaluation of *any* set of options" (Louise 2006, 65-66, emphasis added). Thus, if the criterion of rightness is cashed out in terms of what is (objectively, impartially, universally) best, Direct Consequentialism says we ought to opt for the *best* actions, *best* motives, *best* rules, and so on – whatever makes things go best overall. Indirect Consequentialism, on the other hand, applies the criterion of rightness in the evaluation of

---

[18] This conclusion amounts to the view Railton (1984, 152) has dubbed "Objective" Consequentialism, which differs from "Subjective" Consequentialism in that it concerns itself with actual outcomes and is not wedded to any particular mode of moral decision making. Whereas Subjective Consequentialism prescribes rigorous Consequentialist calculus for any moral agent facing a choice, Objective Consequentialism prescribes whichever method of moral thinking promotes the best result.

only *one* central evaluative focal point, subsequently evaluating every other focal point in terms of their relationship to the central one (Louise 2006, 69). An Indirect *Rule* Consequentialist, then, would deem an act the right thing to do if it fell under the best applicable rule, even if the consequences of that particular act were not the best on that occasion. Indirect *Motive* Consequentialism, on the other hand, condones any act which is motivated by the best motive, and so on.

The argument for Direct Consequentialism[19] owes some of its plausibility to the argument for Actual Consequentialism. Consider this objection to Indirect Rule Consequentialism: Even if we expect some rule to yield the best result *if* generally accepted and followed, there may well occur situations where following some otherwise optimific moral rule becomes a less-than-optimal or decidedly negative alternative, for example due to widespread non-compliance (cf. GENIUS DESERTER above). Thus, even if one might normally expect the rule to yield the best result, it *actually* doesn't in such cases. The agent might achieve nothing of value by sticking to the rule under these non-ideal conditions (Arneson 2005, 239). Arneson (ibid., 240-241) argues that this possibility persists even if one supplements the first-order ideal moral code with nested rules of conduct for non-ideal situations, because one can always find counterexamples where the actual situation is otherwise than postulated in any higher-order ideal rule. If Arneson is right, applying the criterion of rightness to ideal rules – or formulating the criterion of rightness in terms of these rules, as Indirect Rule Consequentialists are wont to do – would always allow for the real-world possibility that some act with suboptimal actual consequences would be the right thing to do. However, if Actual Consequentialism is true, as I have argued, what is *in fact* the right thing to do in a particular situation, here and now, *cannot* be cashed out in terms of such ideal rules or other moral considerations of a conditional nature, relying as they do on evaluations of counterfactual states of affairs. Actual Consequentialism demands recognition of *actual* consequences. This counts in favor of Direct Consequentialism.

Similarly, Parfit (cited in Louise 2006, 69) gives the example of Clare, whose strong love for her child gives her the best motive to have, but who is thereby motivated to give her child

---

[19] For the sake of simplicity, I will in what follows limit myself to arguing for a kind of purely *deontic* Direct Consequentialism, concerned simply with questions of what we ought morally to do, even if I am partial to so called "'global' consequentialist theories", which deem anything affecting the amount of moral value in the cosmos as warranting proper evaluative concern (Louise 2006; Jamieson 2007, 170).

some minor benefit rather than providing some stranger with a larger benefit, and does so. Thus, in terms of actual consequences, Clare acts wrongly, but from the right motive; we would be much worse off if mothers weren't *generally* motivated to take extra care of their children, but total moral value would be maximized if Clare had acted contrary to her otherwise legitimate motive in this *particular* instance. Some theorists take this evaluative conflict – that one can act wrongly from the best motive – to be a decisive argument *against* Direct Consequentialism, because of the seeming paradox of it. However, these objections might owe their plausibility to getting the concept "wrong" confused with "blameworthy" (we have already seen how these concepts come apart) or to the acceptance of some implausibly strong deterministic account of the relationship between motives and acts (ibid.). Any Consequentialist who is *really* concerned with the actual amount of moral value in the universe, must – when considering what is the right thing to *do* – look to the moral value generated by that act, whether or not it is spurred by some otherwise good motive or rule (or disposition, or virtue, or...), or concede that the *actual* amount of moral value in the universe was not, after all, the ultimate object of theoretical concern.

3.4.3 Maximizing Consequentialism

Are we always morally required to do what will yield the best possible outcome, or could we sometimes be morally permitted to bring about some suboptimal possible state of affairs? Maximizing Consequentialism entails the former, and I shall argue in favor of it. The alternative view, Satisficing Consequentialism, contends that "an act X is permissible [iff] its outcome is good enough" relative to some threshold (Portmore 2009). This suggestion has been made to counter some common objections to Consequentialism, namely that it is too demanding and does not leave much room for optional, supererogatory acts (Sinnott-Armstrong 2015, 30).

However, Maximizing Consequentialism might be an attractive *response* to a possible objection to Consequentialism as well, namely the worry that Consequentialism sets too few constraints on what is permissible. If, for example, slavery turned out to be an institution which maximized happiness, even to the detriment of a minority, a Consequentialist theory might allow or mandate slavery. To see why Maximizing Consequentialism might ease such worries, at least on some plausible assumptions, consider the possibility of

SLAVERY WORLD, where some ruling class keeps the majority of the population as slaves. However, the slaves are fairly satisfied with their lot. The rulers treat them kindly; they are fed, clothed and kept safe from harm. In fact, they are happier on average than most people in the real world. What the slaves lack is their freedom, and only their freedom.

Now consider

FREE WORLD, where there is no slavery, and the majority of the population keep themselves fed, clothed and safe from harm. Furthermore, their particular modes of production and organization have made the Freeworlders happier on average than most people in the real world as well. They lack for nothing.

Arguably, Maximizing Consequentialism would value FREE WORLD higher than SLAVERY WORLD, even if happiness-levels were almost identical. Due to the added moral value entailed by the presence of freedom, the former is the optimal alternative.[20] Maximizing Consequentialism instructs us to bring about whichever possible world contains maximal value. It's quite improbable that any world containing slavery would make the shortlist for such possible worlds. There are, conceivably, many alternative modes of organization – at least for humans – capable of yielding high levels of utility without having to rely on slavery. Such states of affairs would be favored by Maximizing Consequentialism. The imperative to *maximize* utility entails a constraint on trying to bring about suboptimal states of affairs, which – in the real world, where it's reasonable to assume that commonsense moral goods such as freedom, fairness, knowledge, love and so on actually have significant moral value in Utilitarian terms as well – entails a duty to establish just social institutions, foster autonomous individuals, enculture loving-kindness, and so on. Thus, Maximizing Consequentialism typically avoids condoning slavery and other perverse institutions, events or states of affairs which might in imaginary examples – and imaginary examples alone – be cast as compatible with maximizing utility.[21]

---

[20] Granted, Maximizing Consequentialism would allow both options if the levels of moral value were *identical*. This, however, would – assuming identical population numbers as well – only be the case because freedom did not affect levels of value. Arguably, one could not object to benign slavery in a possible world where lack of freedom did not affect happiness levels or make a dent in the outcome in any way, but such is not the real world.
[21] As Hare (1981) observes, critics of Utilitarianism will often conjure fantastical imaginary examples to highlight counterintuitive implications of the theory, such as it not being compatible with democracy or freedom, instead condoning tyranny or slavery, on some farfetched assumptions. We should, however, not be

Satisficing Consequentialism, however, would have to permit both possible worlds – slavery or no – insofar as the minimum happiness requirements are met. Not only that, but, as Bradley (2006, 103) demonstrates, Satisficing Consequentialism permits the gratuitous *prevention* of a better outcome by bringing about worse consequences than would otherwise have occurred:

> Let [the number *n* represent the minimum amount of moral value required, and let *n*] be 20. Suppose that, were I simply to mind my own business and continue sitting on my couch, there would be consequences with intrinsic value of +100. Alternatively, I could get off my couch and undertake a course of action to prevent that outcome. This would involve bringing about a different outcome, with intrinsic value of +20.

Satisficing Consequentialism would permit both alternative courses of action, yet it seems clearly wrong to intervene to cause the less valuable outcome when doing nothing would allow a much more valuable outcome (ibid.). Some versions of the Satisficing view might even permit gratuitous killing or torture, so long as levels of moral value do not drop below the threshold (ibid., 102). This seems reason enough to discard such views.

Satisficing Consequentialists are typically out to limit the *demandingness* of morality. But why should we? As Lazari-Radek and Singer (2017, 76) note, demandingness is not intrinsic to the Maximizing Consequentialist view, but contingent on the state of the world. If *everyone* were materially satisfied and otherwise happy, and people's lifestyles were possible to maintain eternally without resources depleting and ecosystems collapsing, and most people actually lead their lives in accordance with the principle of utility, the Maximizing view would *not* demand as much of the agent as it does in our current, imperfect world. Tough luck! Even if aligning with the demands of any true moral theory is difficult, that does not imply that the theory is false.

One criticism aimed at Maximizing Consequentialism from some environmental philosophers is that the maximizing ethos spurs us to consume rather than conserve, which leads to

---

surprised that common intuitions – resulting as they do from our being brought up in the real world – act up when having to deal with fantastical examples. The moral intuitions of an omniscient archangel, operating at the level of critical thinking, might expectably be quite different from ours (cf. section 3.2). Furthermore, if it were really true that humanity – incredibly – *would* prosper under tyranny or slavery, then perhaps commonsense morality should revise its views after all (Hare 1981, 167).

environmental degradation. This, of course, misses the point of the requirement to maximize good outcomes. As Jamieson (2007, 164) puts it:

> [...] any objection that reduces to the claim that utilitarianism requires us to do what is not best, or even good, cannot be successful. Any act or policy that produces less than optimal consequences fails to satisfy the principle of utility. Any theory that commands us to perform such acts cannot be utilitarian.

Whatever the good consists in *is* what ought to be maximized. Thus, if a healthy environment is indeed causative and/or constitutive of the good, or conducive to maximizing it, Maximizing Consequentialism *must* condone it – as well as whatever actions, motives or virtues are necessary to bring it about.

Both Actual, Direct and Maximizing Consequentialism showcase the significance of what Jamieson (2007, 172-173) calls *non-complacency* within Consequentialist thought. The conscientious Consequentialist will always strive to do *better*; she would in some sense regret having failed to do better in one-off situations where that goal could only be achieved through acting contrary to some normally optimific rule or motive, or in situations where doing what was expectably a better outcome would preclude the objectively best, yet less probable, outcome (or, frankly, in situations where the best outcome was simply not achieved for whatever reason). Non-complacency – even if sometimes a psychological burden – seems to me an attractive feature of Consequentialism. It not only entails a certain reasonable sensitivity to the fact that circumstances change, but an imperative to always seek moral improvement as well – a motivating force the absence of which might have left humanity stranded within the confines of institutionalized racism, sexism and other suspect "-isms".

3.4.4 Hedonism

Speaking of infamous "-isms", I want to defend *Hedonism* as the best candidate theory of what is of intrinsic moral value, i.e., what is good in itself ("*the good*"). So far, I have argued that we ought to directly maximize actually good consequences. Now, perhaps controversially, I will argue that the relevant good consequences are *pleasant experiences* (or simply *pleasure*). Other candidate goods – such as freedom, knowledge, beauty, truth, love, ecosystem integrity, etc. – are, I claim, only morally valuable insofar as – and to the extent that – they contribute to making life experientially better for beings capable of subjective experience (Lazari-Radek and Singer 2017, 58). On the flip side, Hedonism also says that the

only thing of negative intrinsic value is *pain* or *suffering* (ibid., 42). Direct (Act) Consequentialism combined with Hedonism yields *Classic Utilitarianism* (Sinnott-Armstrong 2015, 4), which evaluates states of affairs in terms of the net sum of pleasures over pains. This is the view I am defending.

Let me (try to) be precise: When I say pleasure is the only thing of intrinsic moral value, by "pleasure" I mean (a certain aspect of) *intrinsically desirable* (or *agreeable*) *mental* (or *experiential*) *states*. By "pain" I mean whichever mental states could only be *extrinsically* desirable, as means to pleasure, being otherwise objects of aversion and repugnance. This definition of pleasure allows a wide range of phenomenologically distinct sensations and emotions to be counted and valued positively from the moral point of view, insofar as their affective valence is indeed positive. Thus, even if Hedonism is cast as a kind of monism about the good, it has a certain pluralist streak; pleasing phenomenal experiences may be instantiated in several qualitatively different ways (e.g., across species).[22] Hedonism is monist in that it identifies the good *solely* in terms of positively valenced mental states or feelings, but these mental states or feelings may take many forms. This could be restated as the claim that many kinds of first-order experiences exhibit the second-order experiential property of being positively valenced – or, put more simply, *feeling enjoyable* (Crisp 2006; Crisp and Kringelbach 2018, 215). To the Hedonist, it is only the aspect of enjoyment that matters to well-being. As Crisp (2006, 628) puts it: "Feeling good as a determinable is not any particular kind of determinate feeling." Still, we can't help but realize when something feels good somehow. Firsthand experience – not some clever definition – might be the only fruitful path to fully grasping the fundamental concept of enjoyment (ibid., 629).

Underlying my affinity for Hedonism is an intuition best drawn out through another imaginary example. Consider the (Chalmerian) possibility of the

---

[22] It might seem like my account of pleasure is *externalist*, in that I reject that there is a *common* pleasurable sensation which is internally discernable in all cases of enjoyment (Crisp and Kringelbach 2018, 212). This, however, is not precisely my view. I agree with the internalist view that there is a certain "feeling tone" common to pleasant experiences (ibid.), I simply think that the feeling tone might be more usefully defined as an "internal, functional state" (Adolphs cited in Fox et al. 2018, 41) rather than being conflated with its particular phenomenology, to allow for the possibility that, say, *Martian* feelings of enjoyment might be qualitatively different from *human* feelings of enjoyment, even if they serve the same purpose and rank the same in the affective hierarchy, so to speak (e.g., in terms of motivational push and pull).

ZOMBIVERSE – a universe in which there is no emotion nor phenomenal experience.[23] The Zombiverse contains life-like (perhaps even human-like) entities who exhibit (to us) seemingly goal-oriented behavior, as well as somatic responses similar to beings of the actual universe who feel pleasure or pain. However, no Zombiverse entity has the capacity for *feeling* pleasure or pain. There are no *qualia* associated with pain- or pleasure-like somatic responses (Jackson 1982). Observable telic behavior is not actually the result of a subject *having* a goal, in the sense that there is no interior subject there to have it. Thus, there is nothing *it is like* for any one entity to have its life-goals thwarted by any other entity (Nagel 1974). There is no subjectivity at all, and, consequently, no intersubjectivity.

Is any act truthfully described as either morally "right" or "wrong" in the ZOMBIVERSE? I don't think so. The entities of the ZOMBIVERSE would be prime examples of Descartes' mere *automata* (Descartes cited in Kirk 2019), just as surely as the living animals and sentient organisms of the actual world are not. *Ex hypothesi,* nothing would matter to ZOMBIVERSE beings; not their autonomy, not their bodily integrity or sensations, not justice, special relationships nor the stability or intrinsic beauty of the ZOMBIVERSE itself. If morality has a foothold in such a world at all, its applicability is conditional on some kind of subjective interiority emerging somehow. Otherwise, moral constraints on behavior would not apply. This, I take it, is the intuitive argument for morality being in a sense *mind-dependent* (Prinz 2007, 47). Even if we agree with Parfit (2011b, 20), as I do, that there are some irreducibly normative reason-involving truths, which would be true even in the ZOMBIVERSE, in the same way that $2 + 2 = 4$ is true in any conceivable world, we might claim that no such irreducibly normative truth could give us reason to take into moral account any "mindless" entity, unless it were involved in some mind-affecting chain of events.

Of course, this intuition – that nothing would be wrong if no one could care – is not on its own a reason to accept Hedonism. However, it does establish (for us who share in it, at least) that what matters morally, including what is intrinsically good, has got to have something to do with mental states or experientiality. Value – whether moral, aesthetic or of some other normative kind – seems to require at least the possibility of being instantiated, actualized or recognized through a *valuer* to be of much significance. The seeming plausibility of G. E. Moore's (1922, 83-84) beautiful world, eternally unseen, being preferable to an equally

---

[23] See Kirk (2019) for more on David Chalmer's philosophical zombies.

inaccessible heap-of-filth world might owe itself to our *imagining* the contrast and being drawn to beauty, as we tend to be. Had we not (even potentially) been around to imagine the contrasting worlds, no evaluation could be made, and it would not matter which world existed, if any. Similarly, it seems manifestly irrelevant what acts the entities of the ZOMBIVERSE perform, and rather silly to insist that they *ought* to do anything, insofar as it wouldn't matter to anyone.

Morality, then, plausibly has to aim at some kind (or kinds) of value which might affect minds – whether human or non-human – positively. Assuming this much is true, consider the following argument of which it might be a starting premise:

First premise: Morality aims at intrinsically positive mind-affecting value as an end in itself and stands against intrinsically negative mind-affecting value (except as a means to intrinsically positive mind-affecting value).

Second premise: A mind is affected positively iff it experiences positive valence and is affected negatively iff it experiences negative valence. Otherwise, it is not affected.[24]

Conclusion: Morality aims at intrinsically positively valenced experiences and stands against intrinsically negatively valenced experiences (except as a means to intrinsically positively valenced experiences).

The conclusion of my argument amounts to Hedonism as I have defined it. But the second premise might be disputed. Wouldn't it be fair to say that a mind could be positively affected without thereby experiencing enjoyment? Case in point: I am arguably gaining knowledge through writing this thesis, possibly even at such times when writing feels like a dreary and exhausting task. Regardless of how I feel about writing, then, my knowledge may increase, which is arguably a positively mind-affecting outcome. Similarly, the cultivation or

---

[24] Some theorists argue that *neutral* affective valence, if it exists (which many doubt), might serve a similar evaluative role as positive affective valence, signaling *approach*, as opposed to avoidance (Gasper, Spencer, and Hu 2019). Anyway, it seems as if being in a neutral state, if possible, must be preferable to experiencing negative affective valence, so I am going to count such experiences on the plus side.

exhibition of many a virtue might feel strenuous to the mind involved, even if the result is a more virtuous character – a positive, no doubt.

We may, however, doubt it. Knowledge, I submit, would *not* be a positively mind-affecting outcome if it were not conceivably a means to positively valenced experiences. Our tendency to positively evaluate knowledge might simply serve an adaptive function. We know, from experience or through some inborn tendency, that being knowledgeable allows one to attain other valued goods, such as social status, security, friendship, and so on.[25] Thus, knowledge takes on a sheen of intrinsic value, even if it is merely of instrumental value as a means to other ends associated with positive valence.[26] This is a variation on the argument by Crisp (2006, 638) and others that, "over time, human beings have developed dispositions and understandings of goods that, though apparently non-hedonistic, are in fact securely based on their capacity for [promoting] enjoyment." Knowing what is objectively true, however, is not an attractive end in isolation. This fact is perhaps most depressingly illustrated by the prevalence of climate science denial: Facts about which narratives are apt in certain social and cultural settings are more valuable to the agent than facts about geophysical reality, thus truths about climate change – especially of the uncomfortable kind – are easily displaced or ignored through subtle psychological and sociological mechanisms (see e.g. Norgaard 2011).

That actual human minds are not typically drawn to knowledge for its own sake does not, given the non-ideal conditions of our existence, entail that we *shouldn't* aim at knowledge as an intrinsic value. Perhaps we would have such aims if we were always fully rational? I don't think so. Suppose some ideally rational agent were given the choice between

(A) learning some distressing but otherwise inconsequential truth about fundamental reality, perhaps that the external world is nothing but a trick conjured by a malignant demon (Descartes 2017, loc. 4510-4518), or that we are all brains in vats (Nagel 2014, 15-16),

---

[25] As Crisp (2006, 637) notes: "[...] goods cited by non-hedonists are goods we often, indeed usually, enjoy."
[26] Some evolutionary theorists (Hoffman 2019) argue that our sense organs and cognitive makeup – our epistemic apparatus – evolved to allow us to assess and seek fitness payoffs in our environment, not to represent the external world to us accurately or veridically. Pleasurable experiences are linked to the brain's reward system, which serves adaptive functions and enhances fitness through rewarding behavior conducive to reproduction and survival (at least in pre-modern times), such as eating nice and fatty foods, having sex, and so on. Fitness payoffs, then, are typically associated with pleasure. Thus, the initial motivating force behind our (futile) quest for knowledge of the external world may have been pleasure.

or

> (B) simply continuing an overall happy life in blissful ignorance of this truth – about which, we may assume, the agent is powerless to do anything – forgetting ever having been given the choice.

It would *not*, it seems, be wrong – all else being equal – to pass up the information offered and choose (B).

Consider also the possibility of a world where knowledge – its attainment and maintenance – *always* were a negatively valenced feature of experience, inducing some noticeable amount of suffering in the knower. Who in their right mind would value knowledge for its own sake under such condition? No one. Rather, one would try to limit the duress of knowing to those absolutely essential bits of knowledge necessary to maintain a decently worthwhile life (if a decent life is even conceivable in such a hellish world). Knowledge may well be conducive to the good, insofar as it contributes to positively valenced experiences, but is not intrinsically good.

Something similar may be said of virtues. If – in some thwarted reality – being virtuous were always a pain, yet morality demanded that all be virtuous, then morality would prescribe pain universally. Ethics would be the domain of masochists. Luckily, being virtuous is *not* always painful. Rather, a virtuous disposition may be associated with second-order experiences of enjoyment, as well as being conducive to enjoyable experiences overall (Singer 2019, loc. 2876-2915). If this were not so, few would be compelled to care about virtue or morality at all, and it's hard to see why they should be. Suffering, in itself, always gives us reasons to avoid or regret it, except as a means to some greater good (Parfit 2011a, 138; 2011b, 564-569). That greater good, we might add, could not be suffering. Something genuinely good should come of virtue. Imagining, as we have done, worlds in which non-hedonic candidate goods (e.g., knowledge, virtue) are inherently coupled with suffering sheds some light on the priority we intuitively afford to the hedonic tone of experiences and lends plausibility to Hedonism.

48

However, the above argument does not establish that enjoyment is the *only* good which is of positive mind-affecting moral value, i.e., that its presence is a necessary *and* sufficient condition for a mind to be affected positively. Enjoyment may merely be a necessary condition, in conjunction with which other goods might add genuine moral value to a state of affairs. The problem for such views, which Crisp (2006, 640) calls "organic whole views", is that enjoyment seems to be conceptually distinct from other candidate goods, seeing as we can without excessive effort imagine anhedonic lives, in which other goods are present but enjoyment is not (ibid.). What, then, might explain that knowledge (or some other alleged good) is indeed adding value in tandem with enjoyment, but not in the absence of enjoyment? "The idea of an organic whole involves a mystery" (ibid.).

Robert Nozick (quoted in Crisp 2006, 635) set up the following famous thought experiment to dispel Hedonism and other mental state theories of well-being:

> THE EXPERIENCE MACHINE. Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain [...] Would you plug in [for life, preprogramming your life's desires]?

Contemplating THE EXPERIENCE MACHINE, we're supposed to realize that we care about more than pleasurable experiences. Nozick anticipates the objection that plugging in would be selfish by assuring us that *everyone* would be able to plug in. The machine would also be properly serviced and remain functional even if everybody did so (Lazari-Radek and Singer 2017, 44). Still, most intuitively balk at the thought of plugging in. This, Nozick argues, is because we value other things than simply how our lives feel from the inside (ibid.). We want our experiences to be authentic. We want to be in control of our lives, not at the mercy of machines. An *autonomous* life in *genuine* reality seems superior to the inauthentic pleasures available through THE EXPERIENCE MACHINE.

Whatever their initial appeal, such intuitions are not knock-down arguments against Hedonism. They may simply be the result of confounding influences, such as *status quo* bias and risk aversion, perhaps combined with mistrust of real-world technology, fear of being isolated from loved ones in the real world, and so on (ibid., 58-59). I, for one, do not share in them. Assuming I could be sure the machine would actually give everyone plugging in the

best possible experiences in life, without negative side effects discernible to anyone, I would plug in. I might even agree to do so if the experience was not the best possible, merely on the whole better than – or at least experientially identical to – real life. Consider the following possibility, noted by Lazari-Radek and Singer (ibid., 60): What if you are plugged into the experience machine already, and your entire life up until now has been, in a sense, an illusion? Would you unplug? Does the possibility that your experiences hitherto have been machine-made really make them less valuable to you? [27] At the very least, staying plugged in seems a feasible option. Fetishizing authenticity, even when in no position to tell the difference from felt experience, seems a costly alternative.

Many have tried to dodge THE EXPERIENCE MACHINE objection by defining the good in terms of *preference satisfaction* instead of pleasure. So called Preference Utilitarianism, however, has major challenges of its own (Lazari-Radek and Singer 2017, 45-51).[28] The reason it avoids THE EXPERIENCE MACHINE worry, for example, is that Preference Utilitarianism counts as intrinsically valuable the satisfaction of what we desire, not the *prima facie* contents of our phenomenal experiences. Thus, if what we desire is to go for a swim *in the real world*, our having the experience of swimming when plugged into some computer might not count. This, however, yields the strange result that our lives may in some sense be improved without our knowing. You may, for example, form a benevolent desire that some stranger you meet should achieve some goal of hers. She might achieve her goal without the news ever reaching you, and you may never encounter her again. Is *your* life improved by this, even if your benevolent desire was fulfilled? Arguably not. Dislodging well-being from phenomenal experience is an odd move.

3.5   Further objections to Utilitarianism

So far, I have argued in favor of the core tenets of what I have called *Act Utilitarianism*, the view that we ought morally to do that which actually maximizes the prevalence of enjoyable

---

[27] Nick Bostrom (2003) has argued that it is not impossible – in fact, it may be very likely – that we are already living in a computer simulation.
[28] A full defense of Act Utilitarianism against Preference Utilitarianism is beyond the scope of this thesis, so I will not say much about preferences here, but see Lazari-Radek and Singer (2017, 45-51) and Parfit (2011a, 58-81).

experiences in the world. As Sidgwick put it in his canonical formulation: "The conduct which, under any given circumstances, is objectively right, is that which will produce the greatest amount of happiness on the whole" (Sidgwick quoted in Crisp 2014, 233). I will now explore some further objections to this view.

3.5.1 Integrity and the separateness of persons

In his seminal critique of Utilitarianism, Bernard Williams (1973, 116-117) formulates the following worry:

> [H]ow can a man, as a utilitarian agent, come to regard as one satisfaction among others, and a dispensable one, a project or attitude round which he has built his life [...]? It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and [...] his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which *his* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity.

The worry is that I may, according to Utilitarianism, be morally required to set aside whatever deep commitments I may have to some project of mine, simply because I find myself in a situation where some *other* agent's project requires assistance or disruption – by me – in order for maximal utility to be brought about. Thus, the right- or wrongness of my actions might entirely depend on the projects of others and, in a wider sense, on the particular causal nexus of which I am (incidentally) a part, a fact which not only adds to the demandingness of Utilitarianism but threatens my integrity as a moral agent – perhaps even as a person. Nagel (1979a, 203) sums up these worrisome features of Utilitarianism as follows:

> [Utilitarianism] requires you to justify the pursuit of your own personal life and interests only as components of the general good, and does not permit reasons for action to end with a reference to what you want or are devoted to. Those considerations are completely encompassed by an impersonal point of view which accords you no special position, unless it can be impersonally justified.

This worry is a variation on a theme: Utilitarians are wont to disregard or ignore the fact that persons matter to themselves – from the inside, so to speak – in a way which makes it

borderline irrational to subsume everyone under some wholly impartial point of view – or so the story goes, anyway. John Rawls (1999, 24) levelled the accusation against Utilitarianism that it "does not take seriously the distinction between persons". Adopting the Sidgwickian point of view of the universe when evaluating states of affairs is analogous to adopting the point of view of *one* ideally rational impartial spectator (e.g., Hare's archangel). Thus, Rawls claims Utilitarians "adopt for society as a whole the principle of rational choice for one man" who is out to "maximize the fulfillment of his system of desires" (ibid., 23-24). Just as it may be rational for any one person to tolerate some pain *now* to prevent greater pain *later*, it may be, from the impartial spectator's point of view, morally justifiable to inflict pain on some individual(s) to prevent greater pain for some *other* individual(s) (Lazari-Radek and Singer 2017, 81). Thus, whether it comes to inflicting pain or disregarding important life projects, Utilitarianism – aggregating, as it does, costs and benefits between separate people – seems willing to prescribe daunting sacrifices and impose significant costs on some individuals to benefit others, if indeed this makes things go best.

Does this proclivity for sacrifice mean Utilitarianism makes the mistake of ignoring the separateness of persons? On one reading of this objection, Utilitarianism is claimed to reduce individuals to "mere receptacles of pleasure and pain" (ibid.), too easily setting aside other considerations regarding the *subjective* elements of the human condition and the value of a life *for* its possessor – all in the name of some impersonal standard of evaluation. However, it's quite compatible with Utilitarianism to view particular individual lives as inherently valuable and significant, as so many loci of subjective experience, even if the Utilitarian cannot avoid making interpersonal comparisons of value and sum up individual pleasures and pains to get an impartial view of the whole. As I have already argued (section 3.4.4), morality is mind-dependent. Experiences of positive affective valence must be experienced by *someone*, and there is significance in being such a one – recognizable to the Utilitarian. Thus, there is, as Lazari-Radek and Singer (2017, 82) write, "something to be regretted when an individual dies or suffers, even if the upshot of that death or suffering is a greater total amount of happiness [...]." Ideally, no one should have to die or suffer, and we may regret that it is so. Nevertheless, no one can count for more than *one*. If the only way to prevent *more people* dying and suffering is by *my* dying and suffering, then, regretfully, that would be best. Again, as Sidgwick (quoted in Lazari-Radek and Singer 2014, 119) observed:

[T]he good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realised [*sic*] in the one case than in the other.

When discussing the fact that Utilitarianism may sometimes yield *unjust* results, requiring that some innocent person be sacrificed, Smart (1973, 71) says:

Even in my most utilitarian moods I am not *happy* about this consequence of utilitarianism. [...] any injustice causes misery and so can be justified only as the lesser of two evils. The fewer the situations in which the utilitarian is forced to choose the lesser of two evils, the better he will be pleased.

Clearly, then, Utilitarians need not be so detached from their humanity as to ignore the brute fact of subjective reality and be unmoved by the tragedy of necessary sacrifice in the name of the lesser evil. Thus, if an argument from the separateness of persons is to provide solid evidence against Utilitarianism, it must make some different claim. Perhaps that the adding up and summing of costs and benefits to separate people is simply never justified? (Lazari-Radek and Singer 2017, 82). As demonstrated by Parfit (2011a, 185-186), this is clearly false. He gives the example

EARTHQUAKE, in which the only way to save person A's life from some slowly collapsing wreckage is to sacrifice person B's leg.

Alternatively, we could save B from having to endure the pain of getting a leg crushed in the wreckage, thereby leaving A to die. Assuming A could otherwise have had many years of happy life, would this be the right thing to do, morally speaking? Probably not. Even if persons are distinct, few will deny that we may sometimes compare the costs imposed on one individual to the resulting benefits enjoyed by someone else, and justifiably choose the optimal alternative. This is not only true in emergency situations, but holds for a wide range of scenarios, e.g., when imposing legal punishment, doing government budgeting or otherwise allocating scarce resources.

A different reading of the objection from separateness of persons is as an objection not to the Utilitarian criterion of rightness, but to Utilitarian *thinking* and *decision-making*. Williams' initial question is about how anyone can come to regard their own projects or attitudes as

dispensable in the name of morality. It may well be absurd, on some plausible assumptions about human psychology, to expect many to conform to such an ideal. To this, however, Utilitarians can heartily agree. I have already argued that Utilitarianism might not typically prescribe Utilitarian calculation as a decision procedure for humans (section 2.4) and should be kept in brackets at the level of intuitive thinking (section 3.2). Williams (1973, 134-135) claims this move – "by which

> [...] [Utilitarianism] retires to the totally transcendental standpoint from which all it demands is that the world should be ordered for the best, and that those dispositions and habits of thought should exist in the world which are for the best, leaving it entirely open whether those are themselves of a distinctively utilitarian kind or not

– would entail the *disappearance* of Utilitarianism, and he thinks "the residual position is not worth calling" Utilitarianism (ibid.). The idea that a moral theory could be *self-effacing* in this way is alien to Williams. Shouldn't moral theory be of practical significance? (Lazari-Radek and Singer 2017, 95) Utilitarianism could, however, be of practical significance even if its only recommendation were that we adopt non-Utilitarian Rules (of thumb). Utilitarianism "works in practice", Lazari-Radek and Singer argue (ibid.), insofar as its recommendations are actually for the best. Furthermore, the argument that Utilitarianism should be bracketed is contingent on humanity's current psychological and sociological predicament. If the relevant moral agents were otherwise – more like archangels, perhaps – brackets would be uncalled for. This contingency, however, has little to with the *veracity* of Utilitarianism as a moral theory. "[T]he truth of a normative theory cannot depend on contingent facts about the present state of the world" (ibid., 96). Utilitarianism may well be true, even if humans can't handle it.[29]

3.5.2 Utility monsters and repugnant conclusions

Robert Nozick (1974, 41) imagined the possibility of *utility monsters:* beings who "get enormously greater gains in utility from any sacrifice of others than these others lose". It counts against Utilitarianism that it must admit that sacrificing however many people "in the monster's maw" would be the right thing to do. As Parfit (1984, 389) noticed, however,

---

[29] Some might ask: Why go to great lengths to elaborate upon a moral theory that humans cannot handle? Suffice it to say, I think humans should strive towards learning which irreducibly normative facts are true, whether or not we will ever know for sure, and whether or not we will manage to align our conduct perfectly with what is right. Archers who aim are more likely to hit their mark.

utility monsters are a "deep impossibility". Simply *imagining* the level of enjoyment such a being must experience to justify the perpetual sacrifice of others is beyond human capacity. Had we been able to imagine it, we might not balk as violently at the seemingly absurd conclusion that we ought to be sacrificed:

> In the imagined presence of such a [god-like] being, our belief in our right to equality with him may begin to waver – just as we do not believe that the lower animals have rights to equality with us (ibid.)

However, Parfit imagines a more plausible "utility monster"-like case. In the context of considering the outcomes of two rates of population growth, bringing about two non-identical populations differing in both size and average quality of life, he formulates a hedonistic version of what he calls

> THE IMPERSONAL TOTAL PRINCIPLE: If other things are equal, the best outcome is the one in which there would be the greatest quantity of happiness – the greatest net sum of happiness minus misery (Parfit 1984, 387).

This should ring true to any Utilitarian, at least those who accept *Total Utilitarianism,* i.e., that the *total* net amount of happiness is what counts toward moral rightness, as opposed to average happiness per individual (Sinnott-Armstrong 2015, 3-4). Some defend versions of the latter, *Average Utilitarian* view, but we should reject such views. If happiness is the only thing of intrinsic moral value, adding more of it makes an outcome proportionally better, other things equal. Maximizing *average* happiness, however, does not necessarily maximize total happiness. Thus, if we care about attaining the largest possible amount of what is intrinsically valuable (i.e., happiness), we must be Total Utilitarians (Huemer 2008, 928; Pressman 2015, 400).[30] Consequently, we must accept THE IMPERSONAL TOTAL PRINCIPLE.

But, as Parfit (1984, 388) discovered, this principle implies what he called

> THE REPUGNANT CONCLUSION: For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.

---

[30] There are additional reasons to reject the average view as well, such as that it implies states of affairs can be worse even if no one is worse off (Lazari-Radek and Singer 2017, 113).

In other words: Even if there were ten billion people on Earth, each living the happiest life possible, a possible world in which, say, *billions of billions of people*
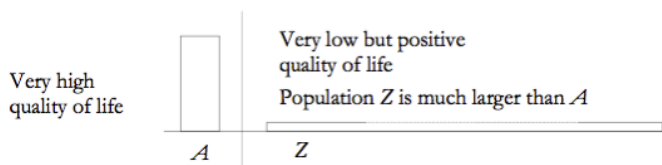

Figure 1

existed, each living a harsh, yet bearable life, would be objectively *better* – if, that is, the sum total of happiness were larger. In other words, "losses in the quality of well-being can be made up for by sufficiently large gains in quantity" (Ryberg 1996, 162). To make the counter-intuitiveness of this conclusion more salient, see Figure 1 (Arrhenius, Ryberg, and Tännsjö 2017). The width of each block represents the size of the population, while the height represents the average quality of life. This is like the "utility monster" scenario, except that the greater amount of happiness to which we must defer is distributed among the vast number of individual lives in Population Z, each barely worth living,[31] rather than being concentrated to one ecstatic, god-like being. It's also a more plausible scenario, and, according to Parfit (1984, 389), easier to imagine. The conclusion, however, may seem just as hard to swallow.

Still, we might bite the bullet, and accept both the repugnant conclusion and other "utility monster" scenarios. We should keep in mind that the vast number of less than optimific individual lives in Population Z *are,* after all, worth living. Initially, we may think of a life "barely worth living" as quite horrible, but, as Ryberg (1996, 167) argues, net welfare in normal, privileged lives may not reach far beyond neutrality most of the time. This may be difficult to establish with any degree of certainty, but it is not a wholly outlandish suggestion. If the individual lives of Population Z are recast as so many normal privileged lives, the repugnant conclusion might seem less repugnant.

Anyway, it should be noted that the repugnant conclusion is not a problem exclusive to Utilitarianism. Every theory which allows that we have at least *some* kind of obligation to make the world better (assuming well-being counts toward betterment) will have to deal with it (Arrhenius, Ryberg, and Tännsjö 2017, 6). So far, no one has found a theory of population

---

[31] We could define a life as being "worth living" if it contains a surplus – however small – of happiness over suffering.

ethics which *avoids* the repugnant conclusion (ibid., 1). *Accepting it*, however, remains a feasible strategy.

3.5.3 The challenge from epistemology

In section 2.4, I admitted that it may be practically impossible for humans to calculate the ultimate results of many (or most) actions of moral significance due to unforeseen future effects. Thus, one might argue, it is impossible for the Utilitarian to *know* what is right or wrong according to their own criterion of rightness (Kagan 1998, 64; Lenman 2000), except perhaps from within the confines of thought experiments where utility levels are simply assumed. The pervasive epistemic inaccessibility of external facts, it seems, issues a daunting challenge to Utilitarianism. Restating the worry in terms of Hare's (1981) two levels of moral thinking, we would simply have to agree that humans are *not* archangels – able to critically assess every relevant nook and cranny of the cosmos – and will likely never be. If so, what role is left for Utilitarianism to play in practice?

As Lenman (2000, 360-361) observes, we might say – from a "disengaged perspective of pure philosophical enquiry" – that the actions with the best consequences are objectively right, even if we do not know which particular actions these are. Lenman, however, thinks this "disengaged consequentialism" can have no practical significance whatsoever with regards to moral deliberation, and thus fails as theory of ethics. This is too pessimistic. While predicting the distant effects of any isolated act will typically be near impossible, we may – for reasons having to do with statistical theory – have greater (but not unshakeable) confidence in our predictions about systemic effects of *patterns* of behavior (Burch-Brown 2014, 111). Thus, the Utilitarian may well be able to give applicable moral guidance in the form of "broadly applied social rules" and "risk management strategies", optimally specified to improve prediction (ibid., 117). A formulation of Utilitarianism which emphasizes Rules (of thumb), then, might be just what the doctor ordered. Furthermore, known unknowns give us reason to be humble in light of our epistemic limitations. Thus, the *disengaged* Utilitarian – while aiming at the same criterion of rightness as other Utilitarians – might be a considerably less self-undermining (and dangerous) Utilitarian than any aspiring human calculator. Being cognizant of the challenge from epistemology, then, helps us arrive at what might be the most viable formulation of Utilitarianism: Bracketed Utilitarianism.

3.6 Bracketed Utilitarianism summarized

In this chapter, I have defended my claim that, at the foundational level (Kagan 1998), Act Utilitarianism may be the theory which yields the most plausible basic moral considerations and specifies the correct criterion of the right. Act Utilitarianism at the foundational level of morality implies Act Utilitarianism at the factoral level as well (ibid.). It does not, however, imply the prescription of Act Utilitarian thinking at the intuitive level (cf. section 3.2). In other words, when enculturing moral norms and values, these need not be Act Utilitarian ones. In fact, there are good reasons why they *shouldn't* be. In most cases, humans are probably not psychologically or epistemically capable of *actually* maximizing happiness by way of Act Utilitarian decision procedures. Instead, we should bracket Act Utilitarianism and enculture some expectably optimific list of moral Rules (of thumb), appealing only to the overall goodness of outcomes when no Rule (of thumb) is applicable or when Rules (of thumb) conflict.[32]

Given the epistemic challenges facing any human agent trying to aim at the best possible outcome *in situ*, moral philosophy should aim at formulating Rules (of thumb) which cover most kinds of agents in most kinds of situations – including emergencies. Philosophers are not archangels, but from the cool comfort of our armchairs we might at least get a little closer to limiting distorting influences on moral thought than possible in the heat of battle, so to speak. Philosophy might provide plausible, state-of-the-art hypotheses concerning which principles should be deployed as part of our moral decision procedures in light of the correct criterion of right action. In what follows, I will try to formulate such hypotheses with regard to the unfolding climate crisis. Specifically, I shall propose some Rules (of thumb) for policymakers wrestling with the politics and ethics of geoengineering.

---

[32] In terms of rights talk, then, Bracketed Utilitarianism may well prescribe a list of optimally specified (i.e., not too complex) *rights* which we have a *pro tanto* duty to respect.

Part II

The Ethics of Geoengineering

Introducing part II

In the above, I have defended a theory of normative ethics. Now, the time has come to apply it. Specifically, assuming the above moral theory is true, I want to explore how or whether it might constrain policymakers in the context of developing and implementing policies for mitigating anthropogenic climate disruption. What are the moral limits on climate policy? I am particularly interested in schemes to deliberately "engineer" the Earth's climate to defend life against the onslaught of global warming, and how such plans might come into conflict with seemingly just rights claims.

Rights talk, as practiced by affected parties in policy debates, might serve as the canary in the coal mine with regard to identifying moral limits. When people invoke rights as trump cards in moral deliberation and discourse, they are making claims about what moral constraints are in play. However, as I argued in chapter 2, legitimate rights claims are the culmination of sound reasoning about what we ought morally to do in terms of more basic, non-rights moral considerations. Rights are not moral primitives. They owe their apparent force to moral reasons which may be formulated without reference to the rights themselves. On my view, whatever force a given rights claim may have, derives from its alignment with Act Utilitarian foundational moral principles. Act Utilitarianism does not afford intrinsic significance to rights, except insofar as they are conducive to the best outcome.[33] The question of moral limits in terms of rights, then, becomes a question of which specific rights claims ought to be respected in a given context to promote the best outcome overall. This, of course, is an empirical question, contingent on many factors. As I have said, we can almost never be entirely certain that we get the answer right on a case-by-case basis (cf. sections 2.4 & 3.5.3).

However, my normative theory prescribes a strategy for moral deliberation which involves *bracketing* Act Utilitarianism most of the time. It is a partly self-effacing theory. Rather than

---

[33] Except perhaps the right to equal consideration (Sinnott-Armstrong 2015, 4). The experiences of every member of the moral universe (e.g., every sentient being) count toward the sum total of value. It follows that any such member has a right to be considered when total value is assessed. Otherwise, we could not be aiming at maximizing total, actual positive value.

require moral agents to try and work out which particular acts are in line with the Act Utilitarian criterion of rightness on the spot, Bracketed Utilitarianism aims at formulating Rules (of thumb). These must be *general* enough to be learnable, applicable to sets of relevantly similar situations and to predictably yield positive outcomes on the whole. They must also be sufficiently *specified* to avoid conflicting with other Rules (of thumb). When, however, such conflict inevitably arises, or when we have to formulate new Rules (of thumb), Direct Utilitarian thinking (aspiring to Hare's *critical* level of moral deliberation) is called for. Not to pin down what is right beyond any doubt (often an impossible task), but to set a course for what is *expectably* best, by identifying which of the available options might correspond to patterns of behavior with optimific systemic effects – whether in terms of available acts, contextually specified rules or, indeed, government policies. That we always try to formulate Rules (of thumb) which are – to the best of our knowledge – conducive to what's *expectably* best, may itself be an optimific Rule (of thumb). While this might strike one as at odds with trying to maximize *actual* good consequences, it may be the best we can hope for (at least until the dawn of some *post-human* era, where technology and science hitherto unimagined have made epistemic barriers to human knowledge of future events all but disappear).

This, then, is what I shall do in what follows: Explore whether there are some rights which may be invoked in opposition to effective climate policies, which ought – as a Rule (of thumb) – to be respected by policymakers in the name of expectably optimific systemic effects.

To limit the scope of my explorations, I shall not consider *every* possible climate policy. I will take as my case the prospect of so called *geoengineering* or *climate engineering* (Robock 2008; Royal Society 2009; Goodell 2010; Keith 2013).[34] Geoengineering is a family name for various technological schemes designed to prevent global warming from spiraling out of control (Boettcher and Schäfer 2017, 266). It is often cast as a last resort – "Plan B" – in the case that human drivers of climate change (most notably: ever accumulating carbon emissions) are not mitigated quickly enough to prevent catastrophic impacts on ecosystems and human societies. Arguably, that is already the case: Much warming is already "locked in", commitments from the parties to the Paris Agreement are lackluster, and there's some

---

[34] I shall be using the term 'geoengineering'.

risk of warming exceeding "climate tipping points", i.e., temperature rise triggering abrupt and irreversible changes, sooner rather than later (Mauritsen and Pincus 2017; Rogelj et al. 2018, 95; Lenton et al. 2019; Tong et al. 2019). It should not come as a surprise, then, that scientists have been seriously contemplating geoengineering for well over a decade. Proposals like fertilizing the ocean with iron to increase oceanic carbon uptake or brightening clouds to increase Earth's reflectivity have been – and are still – vigorously debated. Now, momentum is gathering – outside academic circles as well, among policymakers and entrepreneurs (Boettcher and Schäfer 2017, 269). It is not unreasonable to expect fervent public debate concerning many concrete proposals or actual attempts within the next decade or two. [35] Most likely, rights and conflicting rights claims will feature prominently in such debates. The following is a preliminary attempt to make sense of them.

[35] But see Anshelm and Hansson (2016) for an argument that the geoengineering idea has been fading.

4 What is geoengineering?

[...] attempting to take control of the Earth's climate as a whole [...] is, surely, the ultimate expression of humankind's technological arrogance. Yet if the alternative is to stand back and watch humanity plunge the Earth into an era of irreversible and hostile climate change, what is one to do?

<div align="right">– Clive Hamilton (2013, 16)</div>

4.1   Types of geoengineering

The Royal Society defines geoengineering as "deliberate large-scale manipulation of the planetary environment to counteract anthropogenic climate change" (2009).[36] It is, however, a contested concept, with no static or coherent referent, uniting many widely different proposals for targeted intervention in the climate system (Boettcher and Schäfer 2017, 267).

Geoengineering schemes usually fall into one of two broad categories: Solar radiation management (SRM) or Carbon dioxide removal (CDR) (Jamieson 2013, 529).[37] Clive Hamilton (2013, 20) summarized the distinction succinctly when he wrote:

> While carbon dioxide removal methods target the source of the malady – too much carbon in the atmosphere – solar radiation management methods target one of its symptoms: too much heat.

Carbon dioxide (CO_2) is a heat-trapping gas, with an "insulating effect making it harder for the Earth to radiate infrared heat" (Keith 2013, 47). In excess amounts, it contributes to global warming by allowing less energy from the sun back into the cosmic void, instead

---

[36] As Jamieson (2013, 528) points out, this definition has the strange result that deliberate interventions with similar effects, but different intentions (i.e., *not* to counteract global warming), would not be defined as "geoengineering" proper. For simplicity, I shall ignore this.

[37] We could also distinguish between *system-altering* schemes and *localized* schemes, in accordance with the scope of the respective intervention (Hamilton 2013a, 20).

keeping it here to cause net temperature rise (NASA 2020). Every degree of warming compared to pre-industrial times will wreak more havoc on natural and human systems (Lynas 2020). Current ecosystems and societies are acclimatized to a significantly less energized Earth system (Ellis 2018, 68). More energy will be disruptive, entailing "mostly negative impacts for people, species and ecosystems" (IPCC 2014, 64). SRM and CDR, respectively, offer different ways of lessening that burden. SRM promises less sunlight for heat-trapping gases to trap, while CDR touts less heat-trapping gases in the atmosphere – particularly, less CO_2, the gas to which we owe most of the climatic changes currently unfolding (IPCC 2018, 42).

Both SRM and CDR can be achieved in a number of ways. Here are some of the more plausible schemes proposed:

4.1.1 Solar radiation management schemes

Solar radiation management aims at modifying the Earth's *albedo*, i.e., altering its reflectivity, thereby cooling the planet.[38] There are, broadly speaking, two common proposals for how to go about it: *modifying clouds* and *faking volcanic eruptions*.

The two following schemes for cloud modification seem most plausible:

(1) *Marine cloud brightening*, which involves increasing the reflectivity of clouds, e.g., seeding clouds with tiny sea-water droplets, which will act as cloud-condensation nuclei, allowing smaller, more reflective droplets to form (Goodell 2010, 168-169)

and

(2) *Cirrus cloud thinning*, which involves removing cirrus clouds (high-level clouds that have a net warming effect on the planet) by infusing the atmosphere with bismuth tri-iodide aerosols, producing larger ice crystals in cirrus clouds, thereby reducing the clouds' lifespan (Hamilton 2013, 57).

---

[38] It should be noted, as critics are wont to point out, that "solar radiation management" is a term coined specifically to defuse worries about "geoengineering" – a scarier term by far (Hamilton 2013, 76). Some simply refer to SRM as "solar geoengineering" (Boettcher and Schäfer 2017). I will use these terms interchangeably.

Imitating volcanic eruptions, on the other hand, is a proposal associated with *one* particular scheme – an idea spurred by global temperature drops observed following huge volcanic events, such as the gargantuan 1991 eruption of Mount Pinatubo (ibid.). The scheme is:

(3) *Stratospheric sulfur injection,* i.e., adding sulfate aerosols to the stratosphere to create a radiative shield between the Earth and the Sun, decreasing average temperatures (ibid.).

Hamilton notes: "Stratospheric aerosol spraying is the archetypal geoengineering technique – it would be easy, effective and cheap, and have the most far-reaching implications for life on Earth" (ibid., 59).[39] Indeed, this is one of the geoengineering schemes which has received the most attention, partly because of the relative ease with which it could be pulled off (Robock 2008, 14; Keith 2013).[40] Possibly, even a single country with access to a lot of sulfur and some high-flying jets could do it unilaterally – and might feel compelled to, given the prospect of severe climate impacts on its population (Wagner and Weitzman 2015, 121).[41]

It is important to note that SRM techniques, even if effective at stalling global *warming*, would not solve other challenges associated with rising carbon emissions. Ocean acidification, for example, would still be a problem, unless SRM schemes were part of a mitigation strategy involving significant reductions in carbon emissions and atmospheric carbon levels (Robock 2008, 15; Keith 2013, 8). There are other risks as well, to which I turn in section 4.3.

4.1.2 Carbon dioxide removal schemes

Removing excess carbon dioxide from the atmosphere, permanently storing it someplace where it will not heat the planet, is another type of geoengineering. The aim is to stabilize greenhouse gas levels in the atmosphere below some safe threshold.[42] There are many

---

[39] SRM will likely not be "cheap" by common standards, but may be cheaper than other geoengineering schemes and runaway climate change (Reynolds, Parker, and Irvine 2016, 565).

[40] It was after writing an article on stratospheric sulfur injection that Nobel laureate Paul Crutzen became widely acknowledged as having lifted the taboo on geoengineering research (Crutzen 2006; Boettcher and Schäfer 2017).

[41] Other SRM techniques, such as deploying physical mirrors or sunshields into orbit (Robock 2008, 14), are more far-fetched. I will not devote further attention to them.

[42] Before the industrial revolution, atmospheric CO_2 levels lay around 280 parts per million (ppm) on average (Lindsey 2020). The safe threshold is by many assumed to be 350 ppm (Jones 2017). The preliminary monthly average level of CO_2 exceeds 409 ppm at the time of writing (NOAA 2020).

possible ways to achieve this, including some which are already widely practiced, and may seem less "engineered", such as afforestation and other land management practices (Cox, Spence, and Pidgeon 2020). Hamilton (2013, 20-50) lists the following methods:

(1) *Ocean fertilization* (or *iron fertilization*): By adding iron to stimulate blooms of phytoplankton, we could accelerate the natural processes by which carbon is sequestered in the deep ocean. The plankton absorb carbon dioxide before they are consumed or die and sink to the bottom through the combined mechanisms of gravity and the marine food web, which amounts to a great "biological pump" (ibid., 25-26). Vast amounts of iron would likely be needed to make a dent in atmospheric CO_2 levels.

(2) *Increasing ocean alkalinity*: By adding lime (calcium oxide) to oceans, we could counter ocean acidification and increase alkalinity, thereby restoring the ocean's ability to absorb carbon dioxide by making sea water colder (CO_2 is more soluble in cold water) and protecting the aforementioned biological pump due to increased alkalinity (acidification harms marine life).

(3) *Enhanced rock weathering*: Similar to (2), this proposal would involve hastening the natural process by which rocks slowly break down through contact with rain. This forms carbonates and results in an alkaline solution which is washed into the sea. In the geoengineering version, "rocks would be crushed and chemically transformed so that the carbon dioxide gas in the air became embedded in an alkaline bicarbonate solution", which would then be poured into the sea, increasing oceanic absorption of carbon dioxide (ibid., 41).

(4) *Bio-geoengineering:* We could intervene in land-based biological processes to extract more CO_2 from the air. Notable schemes include:

    a. *Bioenergy with carbon capture and storage* (BECCS), which would involve harvesting biomass (trees, crops, agricultural wastes, etc.), burning it and sequestering those emissions.

b. Producing *biochar*, which fixes carbon through pyrolysis and could then be added to soils to enhance agricultural productivity.

c. *Algal aquaculture* for biofuel (with sequestration).

Of these, BECCS may be the most notable. According to the climate models assessed by the IPCC, many – if not most – mitigation scenarios in which severe or catastrophic global warming is averted involves large-scale deployment of BECCS in the second half of this century (IPCC 2014, 12). This would require vast amounts of land: estimates range from one to five times the terrestrial area of India, depending on the model (Fjellberg 2018).

(5) *Direct air capture*: Drawing CO_2 directly from the air through large-scale industrial processes.

Crucially, none of the CDR proposals currently on the table could be deployed as an *emergency response* to abrupt climate change events. CDR processes are too slow. (Hamilton 2013, 39). SRM, however, could potentially be deployed very quickly. Schemes like stratospheric sulfur injection would require no radically new technology, possibly only modifying commercial air-liner jets already in stock (Keith 2013, 5-7). We'd also need large amounts of sulfur. As it happens, sulfur is an abundant element – there's enough to make two extra moons (Pappas 2017).

It must be noted, as the IPCC recently reported, that "*all* pathways that limit global warming to 1.5°C […] project the use of [CDR] on the order of 100–1000 $GtCO_2$ over the 21st century" (IPCC 2018, 17, emphasis added). CDR, then, has become something of a mainstream proposal within the policy space, implicitly assumed as part of the Paris Agreement (according to which the goal is to "pursue efforts" to limit warming to 1.5°C). We can't, however, expect this to be widely understood among the public (Cox, Spence, and Pidgeon 2020, 744).

4.2   The promise of geoengineering

4.2.1 Prometheans and Soterians

Why would anyone in their right mind suggest geoengineering? Many (if not most) of the proposed schemes presented above have an air of science fiction. Even if they can't all be written off as downright crazy, they are all fraught with complexities, risks and unknowns. Geoengineering has also caught the enthusiastic attention of some climate denialist groups, which might inspire caution (Hamilton 2013, 90).[43]

Hamilton (2013, 123-124) makes a distinction between *Prometheans* and *Soterians*. He chiefly applies this dichotomy to demarcate the worldview of geoengineering proponents from that of its opponents. Named after Prometheus, the overreaching, defiant Titan god of Greek mythology, Prometheans champion anthropocentric, rationalistic, nature-dominating values and "technological thinking" (ibid., 200). They are prone to human hubris and tend to gravitate towards "technofixes", such as geoengineering. Soterians, on the other hand, are namesakes of Soteria, the Greek goddess of safety and deliverance from harm. Hamilton (ibid.) construes the Soterian worldview as one of humility, caution and deference to nature. Soterians are skeptical of meddling with natural processes. Consequently, they tend to oppose geoengineering.

As it happens, the Promethean-Soterian dichotomy can also be applied to two distinct types of geoengineering *proponent*. There's the Promethean proponent: keener on modifying Earth in order to accommodate modern, consumer-capitalist lifestyles than on conforming to its limits.[44] Then there's the Soterian proponent, who reluctantly concludes that catastrophic climate change might already be unavoidable, lest we try to tame or appease Gaia, the mighty beast we have awoken.

---

[43] It also highlights that climate denialism is more about preserving comfortable worldviews than genuine disagreement about facts. The denialist need not be in favor of geoengineering, if he is right that climate change isn't really a problem. He *is* in favor of it, though, because geoengineering seems like a solution more compatible with an attractive, carbon-intensive way of life. For want of a silver bullet solution, it's easier to deny the problem. When a shiny solution appears, however, denialism is superfluous, and possibly no longer worth the social cost. Note, however, that some climate denialists find the idea of geoengineering outrageous (Reynolds, Parker, and Irvine 2016).

[44] The Promethean *opponent* of geoengineering would either have to be skeptical of the practicability of geoengineering schemes or a denier of climate science.

The Soterian sees geoengineering as "a regrettable measure to protect those deeper values now threatened by the consequences of endless expansion – viable societies, vulnerable communities, ecological values and life itself" (Hamilton 2013, 208). Soterian geoengineering proponents are humble in the face of uncertainty. They do not brush aside worries of unintended consequences. Their caution and concern for living things is what makes them even consider such daring ventures. They view Mother Nature not *only* as a caring parent, but as a potentially brutal opponent as well. Now, much to our detriment, we have provoked her. Climate change, to the Soterian, looms as a threat at least as large as geoengineering gone awry. In a sense, climate change *is* geoengineering gone awry: Even though the "engineering" has not been intended, humans have been experimenting with the chemical balance of the atmosphere for centuries through burning fossil fuels, which has put the system out of whack. The Soterian fears it may be too late to undo this damage by way of standard mitigation efforts, such as simply reducing emissions. Either that, or humans simply won't get around to it quickly enough, for reasons having to do with sociology or psychology. Thus, "Plan B" seems the lesser of two evils. If the Soterian is forced to choose between geoengineering or runaway climate change, then geoengineering it is.

4.2.2 Arguments for geoengineering

Soterian proponents of geoengineering must have arrived at their reluctantly supportive conclusion at least partly through considering its promise. What good might geoengineering achieve? Why could it be the better option? Below are some common suggestions. (I consider risks in section 4.3.)

Buying time

Geoengineering may give human societies, animal populations and natural ecosystems more time to transform and adapt. Natural systems and habitats are already changing rapidly due to global warming. The *rate* of climate change is more dangerous to life on Earth than the extent of it (Keith 2013, 26-27). Life tends to find a way, but if changes occur too quickly for animals and plants to be able to adapt, whether by migrating or through the slow grind of evolution, population collapse ensues. Something similar holds for human societies as well. We are a flexible species, but climate change demands rapid transformation of social and technological systems on unprecedented scales, for both mitigation and adaptation purposes.

We have to ween ourselves off fossil fuels and scale back on the expansive and extractive economic activities currently reshaping and depleting the natural world. Otherwise, depending on how much sea-level rise and how many natural disasters these activities end up causing, we may have to relocate significant portions of the global population (see e.g. Risse 2009).

Progress on transforming the global economy in light of this challenge has been dismally slow so far. Fourteen years ago, Paul Crutzen – in his seminal call to arms for solar geoengineering research – lamented the "grossly unsuccessful" attempts at lowering global carbon emissions, noting that annual emissions were still rising, despite the need for a 60-80% reduction to stabilize carbon levels (Crutzen 2006, 211-212). Since then, global emissions have continued to grow – even if some regions and countries report significant reductions (Olivier and Peters 2020). Despite notable political progress (e.g., the Paris Agreement), as well as riveting "green" technological developments and implementation (e.g., solar energy, electrical vehicles, etc.), humanity is way behind schedule. Our current course is set for around 2.9-3.5°C warming above pre-industrial levels, possibly more (Climate Action Tracker 2020; UNEP 2020). Geoengineering, if it can be undertaken safely and effectively enough, holds the promise of more time to cope, for humans and non-humans alike.

Limiting regional climate risks

Climate change is global, but its impacts are local (Keith 2013, 51-52). Consequently, as David Keith has argued (ibid.), geoengineering could probably be tailored to modify both temperature and precipitation on a region-by-region basis. This might be used to limit regional climate impacts, e.g., by reducing the level of precipitation where global warming would otherwise cause increases and *vice versa*. Localized solar geoengineering might even prevent the Arctic from heating so much that its ice sheets melt away and permafrost thaws (Nalam, Bala, and Modak 2018), which – in addition to preserving recognizable lifeforms and ways of life in Arctic regions – could have positive impacts globally as well (see "Steering clear of tipping points" below).

Steering clear of tipping points

Climate scientists have long warned of climate *tipping points*, i.e., global or regional climate shifting from a stable state to some qualitatively different *modus operandi* (Lenton et al. 2008; IPCC 2014, 128). Beyond a certain point, natural processes and feedback loops may cause continued, lasting changes to the Earth system, even if humans no longer exert significant influence on the level of greenhouse gases or other drivers. Surpassing such critical thresholds could lead to abrupt, cascading and irreversible climate disruption. Possible tipping points include:

- Sea-ice melt and ice sheet collapse, accelerating global warming through decreasing Earth's reflectivity (Wunderling et al. 2020) and possibly committing future generations to dealing with sea-levels up to 10 meters higher than today (Lenton et al. 2019, 592).
- Natural carbon sinks becoming carbon sources, e.g., through deforestation and destabilization of the Amazon and subarctic boreal forests or permafrost thawing (ibid., 593-594).
- Changes in ocean and atmospheric circulation, such as slowdown of the Atlantic Meridional Overturning Circulation (AMOC), possibly causing cascades of tipping points around the globe (ibid., 596).

There is uncertainty regarding the level of atmospheric CO_2 which will take us past some tipping point, but recent research indicates such events may be more likely than hitherto believed, risks being higher at lower global average temperature rise (IPCC 2018, 257-258; Lenton et al. 2019). Geoengineering, however, might prevent this from happening, and might otherwise counter some of the adverse consequences.

Protecting the vulnerable

A line of argument related to the above suggestions, but often put forward as an argument in its own right, is the idea that geoengineering might be necessary to protect the world's poorest and most vulnerable groups. Climate impacts are local, and – as a glaring testament to just how little the cosmos cares about justice – the developing nations of the global South,

least responsible for historical carbon emissions, are projected to bear the brunt of the burden of global warming this century (Roy 2018, 2). Economically and socially disadvantaged people will likely have a harder time coping with extreme climate events and losses in agricultural productivity (Keith 2013, 9). While disadvantaged communities in countries at *all* levels of development face more suffering, low-income countries are already less robust in terms of physical infrastructure and effective social institutions. Added climate-related stressors may well fuel increased tension and conflict, as well as spur mass migration (IPCC 2014, 64-73). Cooling the planet, then, might spare the poor and vulnerable from the worst consequences of global warming.

It's just the better option

Another argument, perhaps less obviously Soterian, is that geoengineering might simply be the best available pathway. Consequences may be worse if we don't do it, either in terms of the stringent policy measures necessary to halt warming or with regard to adverse impacts of warming we cannot now avoid. As Briggle (2018, 188-189) puts it: If the gravest predictions of climate science are true, and if the Paris Agreement is not enough, then this is an emergency. *Too much* precaution with regard to geoengineering is possible; we may run the risk of conflating "do no harm" with inaction, and "inaction can postpone or even prevent a better future, which is to say it would be unethical" (ibid., 188).[45] In Utilitarian terms: We should try to minimize damages from global warming in the optimal way, maximizing utility. Regretfully, we may have gotten ourselves into a situation where geoengineering has become indispensable as part of the optimific mitigation package; given that we didn't pursue the best course of action back in the '90s, the *new* best course of action may involve deliberately tinkering with Earth's thermostat.

---

[45] It would certainly be unethical within a Utilitarian normative framework, but plausibly also within deontological frameworks, insofar as they don't prescribe absolute constraints against geoengineering (e.g., on grounds of non-interventionist deference to nature) and the likely harm of refraining from geoengineering exceeds some (arbitrary) threshold providing decisive reason for action.

4.3 Risks

No wonder geoengineering seems enticing given our current predicament, even to some Soterians; successfully executed, it could benefit life on Earth significantly. The alternative might be unacceptably severe climate impacts. What, then, are the *cons*? Many candidate risks present themselves – some pertaining to the nuts and bolts of geoengineering itself, others arising from the complex socio-political, economic and ethical context of its design and implementation:

Practical feasibility

No geoengineering scheme has yet been undertaken at scale. Many schemes, such as liming the ocean or enhanced weathering, would require huge industrial infrastructures and vast amounts of energy (Royal Society 2009, 9). Assuming the necessary resources could be mobilized in time, there's a real possibility that many schemes would not work. Still, the Royal Society (2009, ix) concluded "geoengineering […] is very likely to be technically possible", but there are many feasibility constraints, including (but not limited to): gaining and maintaining political support; getting the science exactly right, ensuring safe and effective deployment and avoiding unintended consequences; access to raw materials, energy and territory; keeping costs manageable. Even if all this could be achieved, there's the legitimate question of why these resources shouldn't simply be directed towards other, less risky climate solutions, like sustainable energy.[46]

Geophysical effects

Many geoengineering skeptics take issue with the glaring hubris of human "mastery projects" (Hamilton 2013, 123-124). Gaia may not be so easily dominated. We don't yet understand nature fully and may likely never grasp all of its complexities. How, then, could we hope to bend it to our will? What the precise geophysical effects of large-scale geoengineering would be is a contentious scientific question at best. The predictive power of the models used to assess effects of untested interventions may not be up to par, given "the difficulty of

---

[46] Whether or not this objection could be successfully answered by the geoengineering proponent, would depend on the scenario under consideration. (See chapter 5.)

adequately representing the complexity of the earth system in models and the abstract nature of the resulting model outputs" (Boettcher and Schäfer 2017, 271).

While there are important, scheme-specific questions of scalability and degrees of invasiveness, CDR methods are in principle better understood than solar geoengineering, because historic records give us a good idea of what the Earth will look like at different levels of atmospheric CO_2. Solar geoengineering schemes, like stratospheric sulfur injection, involve new, additional risks, about which there is considerable scientific uncertainty (Royal Society 2009, x). There are, for example, questions of adverse effects on the ozone layer, ultraviolet radiation and regional precipitation patterns (Boettcher and Schäfer 2017, 268). Many skeptics worry that extensive solar geoengineering may disrupt the monsoon (Robock 2008, 15) and cause droughts (Keith 2013, 52-53), ultimately damaging livelihoods and food supply. Proponents, however, call for a more nuanced debate, stressing that the effects of solar geoengineering would be multi-dimensional, and that outcomes would vary depending on the speed and scale of deployment, which would be a choice (Reynolds, Parker, and Irvine 2016, 565).

Dependency

Solar geoengineering raises another vital issue, the so called "termination problem" (Royal Society 2009, 24; Hamilton 2013, 64): If we manage to avoid dangerous climatic change by way of extensive SRM, without at the same time reducing emissions and atmospheric levels of carbon, we might have to keep engineering the climate for a very long time to avoid a sudden shock to the system. Removing the radiative shield means welcoming the full force of the sun. Should our radiative shield ever fail or be abruptly stopped whilst concentrations of atmospheric CO_2 were still high, the planet would soon return to its pre-SRM temperature levels (or worse). If large-scale SRM is implemented to exert a "high degree of cooling", without CDR and emission reductions, we will become dependent upon the intervention not stopping suddenly.[47] Note, however, that this "termination shock" could probably be avoided if SRM were "slowly ramped down over decades" (Reynolds, Parker, and Irvine 2016, 563-564). These considerations have led many to argue that SRM must never be used, except as

---

[47] Given that CDR methods would not have instant large-scale effects, the termination problem is not as pressing for CDR. We would, however, depend on the long-term effectiveness of the sequestration method used. If large quantities of CO_2 escapes back into the atmosphere because of storage failure, troubles abound.

74

part of a broader mitigation strategy involving emission cuts and/or CDR (Royal Society 2009, x). There is, however, reason to fear that such a comprehensive mitigation plan might be a hard sell in some quarters, considering the intuitive appeal SRM could have for policymakers as a way of avoiding climate impacts *without* necessarily decarbonizing the economy (Jamieson 2013, 530; see also section 5.4.3).

Governance

Geoengineering will have global effects, affecting everyone on the planet. All kinds of successful geoengineering would affect weather patterns and the global mean temperature, while some of the more contentious interventions, like solar geoengineering and ocean fertilization, would bring additional effects of their own, e.g., altering the chemical makeup of the ocean or atmosphere in novel ways. How, then, should geoengineering be governed? Who should – or could – control it? These are still early days, and without better regulatory oversight, there's a real risk that "patents owned by private companies and individuals […] become the *de facto* form of governance of geoengineering" (Hamilton 2013, 80). Already, there are examples of entrepreneurs wanting to undertake audacious ocean fertilization experiments in international waters, bypassing "pesky bureaucracies" (Goodell 2010, 144-162). Thus, government control is called for. However, the inherently global nature of many geoengineering schemes adds to the challenge. As the Royal Society (2009, xi) observes:

> It would be highly undesirable for geoengineering methods that involve activities or effects (other than simply the removal of greenhouse gases from the atmosphere) that extend beyond national boundaries to be subject to large-scale research or deployment before appropriate governance mechanisms are in place.

Yet, given the combination of diverging national interests and unevenly distributed regional impacts from both climate change and geoengineering, is international agreement about geoengineering development and deployment even feasible? Could the world unite over how to set the thermostat? It's not hard to imagine this would be difficult (Jamieson 2013, 532-533), which leads to the next two worries:

Unilateral action

Some geoengineering schemes have been cast by commentators as so inexpensive and technically straight-forward that they might be undertaken by even a single nation, company or wealthy individual, let alone a coalition of like-minded states. Most notable in this category is stratospheric sulfur injection. Many possible scenarios have been put forward: Developing nations who tire of political inertia and take matters into their own hands, desperate to defend their population from the heat (Goodell 2010, 196); well-meaning billionaires wanting to save the planet, teaming up with some heat-stricken country (perhaps in need of investments) to deploy untested technology at large scale (ibid.; Wagner and Weitzman 2015, 116-120); China, with a history of mega-engineering projects, facing environmental protests and climate-induced food-shortages at home, deciding to take action to save the Himalayan glaciers, an important freshwater source (Goodell 2010, 195-196; Moore et al. 2016). Could we also see unilateral *counter-geoengineering* schemes, if some actors on the international stage do not appreciate steps taken by others? Perhaps someone *wants* the Arctic to melt, to be able to exploit otherwise inaccessible resources? (Goodell 2010, 195-196) None of these speculative futures may come to pass. Parson (2014) argues the risk of unilateral geoengineering by non-state actors or small nations has been overstated, and that some powerful state would be the most likely culprit. Nevertheless, the potential for unilateral action (with massive ramifications) remains present.

Conflict and militarization

Unilateral action of the magnitude here considered would likely cause geopolitical tensions to rise, exacerbating political polarization, distrust and the potential for armed conflict and war. Furthermore, geoengineering techniques might be weaponized and used to exert control on the international stage. The military has been interested in the potential for weather manipulation for a long time, and it's hard to imagine that any large-scale geoengineering project would be undertaken without military involvement at some level (Goodell 2010, 206-208). Thus, geoengineering could easily become a part of at least some armed force's arsenal. The temptation to use it for military purposes might prove a greater motivating force than precaution.

76

Legal issues

There is already much talk of how costs and damages from anthropogenic climate change might call for "climate reparations"; high emitters owing compensation to those harmed by global warming (Mark 2018). While this may seem *prima facie* just, it's contentious whether global warming, being an *unintended* consequence of utilizing certain forms of energy, could really be blamed on emitters as such – at least not retroactively. Thus, the case for climate reparations is not straightforward. In a future *geoengineered* world, however, every kind of weather would become someone's fault (Goodell 2010, 205). Legal ramifications could be massive. (Find your house flooded after heavy rain? Sue the government! Caught in a blizzard? Sue the government!)

Psychological effects

Geoengineering might change our environment quite palpably. For one, injecting aerosols into the stratosphere will change the color of the sky, giving it a "white, cloudy appearance", as well as yielding more dramatic sunsets, such as the one caused by the 1883 Krakatau eruption in Indonesia, which inspired Edvard Munch's *The Scream* (Robock 2008, 16). How will the loss of blue skies make us feel? For many, geoengineering might represent the end of wilderness and the final termination of humanity's historically subservient and co-dependent relationship with nature. This may foster what has been termed "ecological grief", as a response to ecological loss (Cunsolo and Ellis 2018). Geoengineering will impact humans and non-humans along many dimensions, in ways not well understood. Its psychological effects may be significant.

Moral hazard

"Moral hazard" is a term used by economists to describe "loss-increasing behavior that arises under insurance" (Rowell and Connelly 2012); risky behavior is more readily undertaken when the agent feels assured of rescue. The argument from moral hazard states that pursuing geoengineering as a course of action – simply researching it or talking about it publicly, even – might delay necessary mitigation or draw resources from climate adaptation by undermining public support for these efforts (Royal Society 2009, 37). Why vote for painful emission cuts, if geoengineering could do the trick? There is widespread agreement among

scientists that geoengineering should not be humanity's sole response to climate change, seeing as it "cannot perfectly compensate for the multidimensional climate changes produced by greenhouse gases" (Keith 2013, 8). Continued emissions would not only exacerbate impacts, but also increase our dependency on hitherto unproven technologies. If this is not sufficiently appreciated by publics, moral hazard becomes hazardous indeed.

These are just *some* of the possible pitfalls of geoengineering. In the next chapter, I will briefly sketch one plausible scenario for the future of geoengineering, before I consider how emergency ethics and rights may play into it.

5 Geoengineering and rights

5.1   A possible future

Although many questions remain unanswered, geoengineering has already entered the collective imagination (Boettcher and Schäfer 2017, 274). Policymakers, academics, entrepreneurs and financiers have been grappling with it for years. The larger public is next.

I'll allow myself some blatant speculation concerning the future. Consider the possibility of

> EMERGENCY GEOENGINEERING: World-leaders recognize that they won't be able to keep warming below 2°C without sacrificing much in terms of short-term economic growth and public support. The required effort necessitates too many draconian measures for most politician's tastes. Thus, geoengineering seems increasingly tempting. Soon, geoengineering is on everybody's lips, as the parties to the United Nations Framework Convention on Climate Change gather to discuss steps towards honoring the Paris Agreement despite insufficient emission cuts.[48]

Many countries are skeptical at first. However, increases in the number and intensity of dangerous climatic events – such as cataclysmic forest fires and droughts – tip the scales in favor of giving geoengineering a try. The first real-world geoengineering proposal put forth consists of a suite of geoengineering interventions to optimize total effects (Rasch, Latham, and Chen 2009, 6; Hamilton 2013, 108-109). CDR methods (like BECCS) take center stage initially, until real-world constraints lay bare their limitations (e.g., tensions concerning land-use). However, solar geoengineering research and testing is an important part of the package from the get-go. As the century progresses and climate impacts become a more pressing concern, there is a surge in calls for emergency measures with immediate effect. At this stage, some capable agent within the policy space – whether the

---

[48] Already, in their next assessment report, due in 2022, the IPCC will address geoengineering *governance*, not merely its technical aspects (Boettcher and Schäfer 2017, 270).

United Nations or some other affected party, perhaps with the tacit support of members of the UN Security Council – attempts a solar geoengineering intervention (likely some kind of stratospheric aerosol injection) to try and stabilize temperatures at a level which will make no nation worse off in economic terms (at least no *powerful* nation).

Assuming the EMERGENCY GEOENGINEERING scenario is plausible, would such geoengineering interventions be morally sound, especially with regard to rights infringement? That's the question I'll be exploring in this chapter, from within the normative framework of Bracketed Utilitarianism. I'll be devoting extra attention to solar geoengineering, as this kind of intervention is thought to highlight the ethical conundrums of geoengineering most starkly (Hamilton 2013, 158). Before I get to the "do's and don'ts" of geoengineering, however, I will discuss some relevant topics of emergency ethics more generally.

5.2 Emergency ethics

5.2.1 Defining the emergency

If EMERGENCY GEOENGINEERING, or some similar scenario, comes to pass, geoengineering interventions will be pursued as emergency measures. We may, with Sorell (2003, 22), define an "emergency" as "a situation, often unforeseen, in which there is a risk of great harm or loss and a need to act immediately or decisively if the loss or harm is to be averted or minimized." Of course, calling the kind of events which could inspire geoengineering interventions "unforeseen" would be a gross misnomer; large-scale climate impacts *have* been foreseen, the moral significance of which has been pointed out by many (see section 5.4.1). The point at which solar geoengineering is undertaken in our example, is the point at which certain projected climate impacts have gone from latent threat to manifest risk, promising even greater harms and losses than those suffered at the hand of global warming until then. We're picturing a point at which climate disruption is no longer simply foreseen, it is *felt* – by enough people to instill in decision-makers an unprecedented sense of urgency, whether from altruistic care or deference to their base of supporters, making them assert "a need to act immediately or decisively" (Sorell 2003, 22). Of course, the precise meaning of "immediate" or "decisive" action is sensitive to context. Our scenario would likely leave more room for cold-headed deliberation concerning countermeasures than would

public emergency scenarios of a more local, less complex nature. Geoengineering decision processes could span weeks, months or years.

Hamilton (2013, 156) issues a challenge to "those who defend solar radiation management research as a form of preparation for a crisis". He claims they

> have yet to provide answers to the following questions: What are the criteria for a climate emergency requiring rapid intervention? Who would determine that an emergency exists? Who would authorize the emergency response, and from where would they derive their legitimacy? Who would decide that the emergency is over?

To the questions of climate emergency criteria, we could reply (Bracketed Utilitarianism in hand): When *refraining* from rapid climate intervention would likely yield more net suffering than any alternative course of action, there is a climate emergency requiring rapid intervention. Depending on our assessment of the risks associated with the intervention in question (i.e., geoengineering), we could qualify this definition with a requirement that more suffering be *very* likely before intervention is required. If we were to stick more closely to Sorell's "emergency" definition, we might also introduce a requirement that the quantity of net suffering (i.e., harm or loss) surpass some threshold making it "great". While this would not be a principled Consequentialist move as such, it might correspond to an optimific Rule (of thumb): to always seek as robust justification for high-stakes interventions as possible (e.g., to increase legitimacy among affected parties, minimize the number of moral mistakes made, etc.).

Conversely, the emergency is over, and further geoengineering uncalled for, when continued intervention likely yields net suffering compared to alternative courses of action. Here, again with reference to the risky nature of geoengineering, it might be prudent to lower the bar, requiring the intervention be discontinued as soon as it seems likely to result in net suffering by however small a margin. We should not keep tinkering with the climate-system for long after decidedly negative consequences have become apparent.

In EMERGENCY GEOENGINEERING, it's assumed that those authorizing the emergency response would be high-level politicians affiliated with the UN – heads of state, or some such. If so, their legitimacy would hinge on that of the respective political processes underpinning their ascent to leadership. In a cosmopolitan Kantian vein, we could assert that

the legitimate political authority of state leaders is provisional and conditional upon their contribution to the full realization of a world order conducive to what Ypi (2012) calls "the principle of right", i.e., the criterion of rightness (see section 3.4.1). Kantians, like Ypi, typically spell out this criterion as "just, reciprocal relations", but Bracketed Utilitarians might substitute that for "expectably best consequences", among which such relations should be counted.[49] Thus, Bracketed Utilitarianism seems capable of answering Hamilton's challenge to the aspiring geoengineer, at least in the context of our EMERGENCY GEOENGINEERING scenario.

5.2.2 Practical reasoning under exceptional circumstances

Let's now imagine ourselves being in the position of one of the policymakers due to decide whether to undertake some geoengineering intervention as an emergency response. For sake of argument, assume we are a legitimate occupant of the room where it happens. Furthermore, ours is the deciding vote. Would Bracketed Utilitarianism be of any use to us as a moral *decision procedure*? Could it guide our practical reasoning?

Recall that the Utilitarian criterion of rightness pertains to a level of critical thinking toward which humans stretch in vain, inherent epistemic limitations barring access to ultimate facts about the objective value resulting from any given intervention (see section 3.5.3). Actual, remote consequences are not visible to us. We could not be certain what the causal ramifications of our solar geoengineering gambit would be, except we'd know they'd be massive and affect the identity of future people (Lenman 2000). That leaves little to go on. Even if we could successfully define the emergency in terms of the relative likelihood of negative outcomes, we could not be 100% confident we'd solve it in keeping with the objective criterion of rightness. It warrants restating: Due to the ever-present possibility of our actions having massive, unforeseen ramifications, contemporary humans cannot hope to *know* if we're acting in accordance with what we ought morally to do from the point of view of the universe. We're doomed to aim for what's expectably best from our own, subjective point of view.[50] Bracketed Utilitarianism aims to formulate Rules (of thumb) conducive to that end.

---

[49] It's not obvious, however, that some world leader *not* committed to ideals of justice would thereby act wrongly in the fact-relative sense by spearheading or supporting some geoengineering intervention. That depends on the actual consequences (cf. Parfit 2011a, 212-232).

[50] Consequently, most of our moral judgements concerning events in the real world should be stated in terms of Parfit's belief- and evidence-relative senses of moral wrongness (see section 3.3.1), except if our judgement is *a*

Some theorists of emergency ethics hold that acts performed during emergencies, or in other extreme cases, should *not* be principled or rule governed. Case in point: Harel and Sharon (2011) argue there are "circumstances in which it is simply wrong to use rule-based reasoning" (ibid., 852). If that is true, formulating emergency Rules (of thumb) seems a pointless endeavor. But is it true?

Harel and Sharon argue from a deontological standpoint (ibid., 845). Their argument amounts to the claim that we should never try to internalize rules containing exception clauses (exceptions being what extreme situations would demand), because that would normalize the exceptions and undermine the general principles of moral law. These, they argue, must be unconditional (i.e., not contingent on context-sensitive facts). Hence, instead of recommending we appeal to some contextually specified rule when deciding what to do in emergency cases (for example when we are in a position to save the many by sacrificing the few), they recommend we keep our moral principles unspoiled – intact as *general* rules (cf. section 2.2) – and simply put them out of our minds, setting them aside as a matter of practical reasoning *in extremis*.

On this view, even if agents facing the relevant kind of moral dilemma are on principle forbidden from sacrificing the few, they should forget about that, and do it anyway, insofar as "brute necessity" demands it (ibid., 862). Necessity will demand it, the argument goes (ibid., 859-862), if you find yourself in a position to save innocent lives, while recognizing that the only way of doing so is by sacrificing someone. So, instead of blindly following the rule against sacrificing someone, you should – under exceptional circumstances – simply act upon your recognition that you have a duty to save innocent lives. Presumably, you would be excused for doing whatever that would take.

Underlying Harel and Sharon's position is a desire to harmonize certain deontological intuitions with the intuition that an agent could not be blamed for sacrificing the few for the many when necessary. They want to have their cake and eat it too, claiming both (1) that one should never sacrifice the innocent and (2) that sacrificing the innocent should sometimes be

---

priori true (e.g., it would not be wrong in the fact-relative sense to withhold from your dying grandmother some bad news on her deathbed, insofar as she could not alleviate the situation and it would only make her last hours worse).

done by someone (ibid., 860-861). Thus, they make an odd move: stating both that we have an overriding moral duty to save innocent lives, and that this duty should not be codified and embedded in our practical reasoning. At all costs, the decision to sacrifice the few must *not* follow from a line of reasoning deploying a principle or rule allowing such acts on certain conditions, because conditionality of any kind would weaken deontological constraints (ibid., 858-859). It looks almost as if they're aiming at some kind of esoteric morality, even though Deontologists typically shun that idea (see section 2.3). Esoteric morality, however, involves keeping the true moral rules secret, whereas Harel and Sharon seemingly want us to forget about rules altogether when necessity calls – except (for reasons somewhat unclear) the rule that one should always try to save innocent people (a tenet defining necessity itself, on this account), which we should adhere to without making it explicit (ibid., 859).

Harel and Sharon's argument is a brave attempt at saving Deontology from itself. In order to counter the criticism that no non-arbitrary deontological principle can deal satisfactorily with emergencies (cf. section 3.3.4), they deemphasize *objective* criteria of moral value in favor of *subjective* criteria: They stress that "within a proper deontological conception, the reasoning of the agent is a key factor in the determination of the moral status of his actions" (ibid., 859). In other words: If there is an emergency at hand, and you must act, you need not heed any particular rule concerning what to do, so long as your heart is in the right place – your intentions aligning neatly with your duty to save innocent lives. This arrangement is meant to defend deontological constraints from being undermined by Consequentialist practical reasoning.

One worry immediately springs to mind, however: Wouldn't the knowledge that acts performed under exceptional circumstances were exempt from general rules likely weaken deontological constraints as well? Or are exceptional circumstances so easily recognized that no mistakes would be made regarding when to disregard deontological constraints? If not, maybe we'd be better off if we had some ultimate moral principle to fall back on, such as the principle of utility, rather than discard principled thinking? We have no guarantee that agents abandoning rule-based reasoning would be compelled to act upon the same duty of beneficence – or interpret it the same way – as Harel and Sharon would like.[51] If we applied

---

[51] We might at least be *marginally* better off with a more principled backup decision-procedure. Of course, the principle of utility – and other rules – could be wrongfully applied as well (see section 2.4).

the principle of utility under exceptional circumstances instead, at least it wouldn't be arbitrary whether we ended up saving the many or the few. This by itself is a compelling reason not to give up on Rules (of thumb).

If I may say so, Harel and Sharon's view looks a little like a deontological inversion of the "Rules (of thumb)" idea – a kind of *Bracketed Deontology* (BD). They want to defend general rules in times of normalcy, while allowing that these rules be bracketed – that is, set aside when we're deciding what to do – in emergencies. They argue we should defer to an overriding duty of beneficence toward innocent people when general rules are thus bracketed. Similarly, on my account of Bracketed Utilitarianism (BU), we should stick to general rules so long as general rules can be successfully applied. If not, we must defer to the principle of utility. One notable difference between the views, however, is the respective contexts in which the bracketing of "true" morality (Deontology and Utilitarianism respectively) is prescribed; BD brackets Deontology in exceptional circumstances (e.g., when there is an emergency), whereas BU brackets Utilitarianism when circumstances are *not* exceptional. Notice that this makes the views quite similar in practice – at least in normal times, when Rules rule supreme. Another difference is that BU does not discourage principled thinking under exceptional circumstances – urging us, as it does, to consider expectable net utility when general rules fail. BD, on the other hand, explicitly discourages principled thinking in emergencies, instead recommending a more automatic decision procedure.

It seems to me that this total reliance on the automatic responses of humans under duress counts against BD. Even if we enculture the correct moral rules for normal circumstances – thereby making them widely shared norms or habits influencing decision-making (under exceptional circumstances as well) – we would be naïve to trust that this would translate into the right decisions being made by *every* agent with the power to save the many by sacrificing the few. Of course, almost any moral philosopher putting forward some guide to action could be accused of similar naiveté (recall, again, the practical impossibility of precise Utilitarian calculation), but it seems that giving up on principles altogether might issue an overly tempting invitation to those bored from – or incapable of – justifying their choices with reference to the best possible moral reasons. Of course, Deontologists (like Harel and Sharon) would take issue with Utilitarians (like me) concerning what those reasons *are*. It is the staunch belief that such reasons could not be Consequentialist that forces them to give up on

principled reasoning in times of crisis in the first place. Yet… isn't that a high price to pay to save Deontology?[52]

To make a long story short: I disagree with the sentiment that "it is simply wrong to use rule-based reasoning" under circumstances of necessity. To the contrary, we need the best possible Rules (of thumb) we can muster, hedging our bets in case significant harm or loss would come of *ad hoc* emergency decision making. (Large-scale solar geoengineering schemes being set in motion due to unprincipled thinking is surely a sobering prospect.)

5.2.3 Rules (of thumb) for geoengineers

What, then, would be the items on the Bracketed Utilitarian's list of likely optimific, sufficiently specified Rules (of thumb) governing emergency choice situations like the one facing our geoengineering-pondering policymaker? Keep in mind that we're dealing here with a *public* emergency, "facing whole states or large numbers of people", which public agencies and their officials have a special obligation to handle (Sorell 2003, 22). I'll brave making two tentative suggestions for Rules (of thumb) pertaining to public emergencies in general, which would be applicable to our case (though formulating a complete list surely requires more work). *Private* emergencies might warrant different Rules (of thumb).

Two Rules (of thumb) for dealing with public emergencies

> R(ot) 1: *Time and resources permitting, assess every relevant piece of information from reliable sources regarding every possible alternative and outcome before deciding on an intervention.*

This follows from the Rule (of thumb) to always seek as robust justification for high-stakes interventions as possible (section 5.2.1). We may be stuck in the murk with regard to ultimate consequences, but the least we can do is *try* to light the epistemic torch. It is, of course, debatable whether knowledge production yields expectably optimific systemic effects *as such*, but it seems reasonable to assume that decision-makers could more easily decide in

---

[52] This is not to say that Bracketed Utilitarianism *always* recommends rigorous principled thinking. Some emergency scenarios may truly leave "no time for philosophy" (Briggle 2018, 188), demanding immediate, automatic responses (fleeing, ducking, etc.). This, however, would not be the case for the kind of public emergency currently under consideration (see section 5.2.1). Furthermore, even automatic responses should ideally flow from (expectably optimific) internalized rules or principles.

favor of the best possible consequences if the foundations of their estimates regarding consequences were as solid as possible. However, the "time and resources permitting" proviso is of paramount importance; too much time spent researching possible outcomes could prove fatal in any emergency. With regard to geoengineering, though, there is likely still sufficient time to research alternative options extensively (Hausfather and Peters 2020).

R(ot) 2: *Always opt for the alternative least likely to cause the permanent extinction of known lifeforms.*

Given that morality is mind-dependent (which I argued in section 3.4.4), there can be no morally relevant value in a world without life. If life is extinguished, a potentially infinite quantity of future enjoyable experiences (or other moral goods) is forfeit. If life is sustained, even just barely, futures containing infinite enjoyment (or other moral goods) are still possible (Parfit 2011b, 612-621). Thus, insofar as mind-dependent moral goods are what's valuable, the overriding moral concern should be to sustain life in the knowable universe. Currently, we know of no other lifeforms than those on Earth. Consequently, sustaining life on Earth should be our overriding moral concern.

5.2.4 The status of rights claims in emergencies

The framing of our EMERGENCY GEOENGINEERING scenario raises questions concerning rights. It's widely thought, at least by people applying the general conception of rights (see section 2.2), that circumstances of emergency may warrant the "turning off" of certain rights, insofar as this may be conducive to restoring normalcy and "circumstances of justice" (Murphy and Whitty 2009, 230; Sandin and Wester 2009, 292; Freeman 2016, 24-25). Bracketed Utilitarianism accommodates this idea in two steps: (1) by aspiring to formulate Rules (of thumb) that are sufficiently specified to handle exceptional circumstances (meaning some rights may be explicitly conditional), and (2) by keeping Utilitarianism at the ready, in brackets, prepared to pull its weight when the Rules no longer suffice (e.g., when rights conflict). It's important, however, that we do not reach for Utilitarianism at the faintest whiff of emergency. I have said that principled, rule-based reasoning should have a role under exceptional circumstances. I also said that rights should be maintained as part of our moral repertoire, because the social institution of rights talk may have expectably optimific systemic effects (this seems probable, anyway, after reflecting upon the course of human

history anterior to the UN's 1948 *Universal Deceleration of Human Rights*). Thus, in order to protect the institution of rights from corrosion or perversion, we cannot simply "turn off" rights under exceptional circumstances, unless circumstances are sufficiently exceptional.

What constitutes sufficiently exceptional circumstances? Doesn't regulating this require another Rule (of thumb)? Yes, indeed. The straight-forward, Utilitarian proposal is similar to the one I gave in section 5.2.1: When *not infringing* on relevant rights would likely yield more net suffering than rights infringement, there are sufficiently exceptional circumstances to justify infringing on relevant rights. However, we must consider the consequences of any given Rule (of thumb) before endorsing it, and there is reason to fear that this formulation might have the unintended effect that legitimate rights are wrongfully infringed, or the institution of rights talk undermined, seeing as it requires very precise calculation. In many cases, agents may simply be incapable of discerning which alternative nets suffering. Differences in projected utility levels may be very small between alternative courses of action, and influential people may be able to leverage this for sinister ends. To minimize the likelihood that our Rule (of thumb) has these perilous effects, we could amend the proposal as follows:

> R(ot) 3: When not infringing on relevant rights would *very likely* yield *significantly* more net suffering than rights infringement, there are sufficiently exceptional circumstances to justify infringing on relevant rights.

At least in the context of geoengineering, it looks *prima facie* improbable that this precautionary formulation would beget disaster. Sidestepping rights to implement geoengineering in a hurry seems the greater risk. It also resonates well with common intuitions about when emergency measures are called for, namely when *great* harms or loss are imminent, which would further bolster the legitimacy of any decision made under this Rule (of thumb). Geoengineering risks would multiply if any given intervention were seen as illegitimate (recall "Governance", section 4.3), so these are prudent precautions. Of course, "very" and "significantly" are vague terms, referencing indeterminate cut-off points on the scales of likelihood and suffering, so we would have to rely on our leaders exhibiting what Aristotle called "*phronesis*" – practical wisdom combined with ethical virtuosity (Kraut 2018) – when operationalizing this Rule (of thumb).

I turn now, finally, to the question of what guidance Bracketed Utilitarianism could provide in the context of assessing particular (possible) rights claims standing against geoengineering. The main proposal under consideration is EMERGENCY GEOENGINEERING, i.e., the rapid, large-scale deployment of solar geoengineering to prevent runaway climate change. (However, some of the following claims will be of relevance when considering CDR schemes as well.) In the spirit of Universal Consequentialism (see section 3.4), I will – in addition to human rights – briefly consider rights claims pertaining to nonhuman entities as well, such as animals and nature itself.

5.3   Rights versus geoengineering

5.3.1 Human rights

The Universal Declaration of Human Rights (UN General Assembly 1948) is a likely source of future rights claims made against geoengineering. Its Article 3 reads: "Everyone has the right to life, liberty and the security of person." Geoengineering could threaten these rights for a significant number of people, for example by spurring climatic events causing famines, conflict and forced relocation. The same, however, goes for runaway climate change. Famine, conflict and mass migration threaten either way. Thus, if some (contemporary or future) affected party asserts their right to life, liberty and security as an argument against geoengineering, and some other (contemporary or future) affected party asserts the same right in a call for geoengineering, decision-makers will be facing conflicting rights claims. If so, Bracketed Utilitarianism bids us aspire to the critical level of thinking (section 3.2), sum up the consequences to the best of our ability and opt for the lesser evil. Of course, whichever option is chosen, governments would be obligated to try and prevent the loss of lives, liberty or security, e.g., by aiding adaptation efforts.

In accordance with R(ot) 3, geoengineering should only be undertaken despite rights infringement if this would very likely prevent a significant net amount of suffering (section 5.2.4). Furthermore, my R(ot) 2 demands we always try to minimize the probability of life's extinction. These considerations might speak (softly) in favor of geoengineering, seeing as runaway climate change could (though it is unlikely) set us back to a level of technological development where humanity would not be capable of averting other existential risks, such as

an asteroid strike, because Earth might enter a "hothouse" state, triggering cascading climate impacts which would make advanced civilization untenable (Lynas 2020, 239-283). In addition to impairing extinction-level event preparedness, this extent of warming would threaten life, liberty and security for people all over the world, whereas areas adversely affected by solar geoengineering could probably be geographically limited (Keith 2013, 9). If so, the net amount of suffering from runaway climate change would be significant compared to that from geoengineering. Geoengineering is less likely to have these cataclysmic effects, given that it could be ramped up or down in a controlled manner (Reynolds, Parker, and Irvine 2016, 563-564) – assuming, that is, that unintended or unforeseen side-effects wouldn't trigger runaway climate change. These, however, are not confident assertions. More research is needed.

Several other rights invokable against geoengineering would involve the same kind of dilemma, where both geoengineering and runaway climate change would pose threats relevant to the rights claim at hand (call this *the Similar Threat dilemma*). These include:

- The right to an adequate standard of living conducive to health and well-being (UN General Assembly 1948), which the geophysical effects of both failed geoengineering and runaway climate change could preclude for many people.
- The right to a social and international order conducive to the realization of other rights (ibid.), which could be hampered by both climate induced conflict and tensions over geoengineering governance or unilateral action.
- Rights to traditional ways of life (Norman 2017), including resource management (e.g., fishing, hunting and harvesting) and cultural practices (e.g., access to sacred sites), which for indigenous groups are already under pressure on many fronts, a development which could be exacerbated by both local effects of geoengineering interventions (e.g., the construction of requisite infrastructure) and climate change.

Other assertable rights claims, particularly pertaining to risks regarding governance, would *not* automatically engender the Similar Threat dilemma. Consider, for example, a sovereign state or nation's right to self-determination. If geoengineering is not pursued unanimously, but unilaterally or by a group of states – even by a majority – some actor on the international stage, feeling left-out of the decision and/or worried about the consequences, might claim

their right to self-determination or sovereignty is infringed, at least if the intervention affects their people and territory (which it likely would, directly or indirectly). Here, decision-makers would have to decide whether to disregard some actor's right to self-determination in the name of other relevant moral considerations, such as future people's right to life, liberty and security. Seeing as self-determination or sovereignty does not seem like a concern able to override R(ot) 2, geoengineering might have the upper hand here.

However, if the resistant actor is some powerful, populous state with aggressive leadership, the scenario becomes fraught with risks. Perhaps there could be a war over geoengineering? If so, we might actually be facing something like the Similar Threat dilemma after all. In the age of nuclear weapons, both large-scale, interstate war and runaway climate change could lead to civilizational collapse. In this scenario, Bracketed Utilitarianism would prescribe attempting to calculate the value of the consequences and aiming for the significantly greater good. In the exceedingly unlikely case that there's a credible threat of nuclear war over solar geoengineering, and this threat cannot otherwise be neutralized, the best course of action is probably to refrain from geoengineering and pursue other measures. After all, we've not yet reached the point where reducing emissions and curbing the consumption of natural resources couldn't even *soften* the blow of climate change. Belated implementation of insufficient, standard mitigation strategies, leading to excessive, yet survivable, global warming, would likely be preferable to nuclear holocaust.

5.3.2 Animal rights and rights in the natural world

Many have argued convincingly that – although humans and animals should not be treated the same in all respects – there is no morally relevant difference between *Homo sapiens* and other, nonhuman animals which could justify our denying them the right to life, reasonable autonomy and freedom from wanton cruelty (e.g., Singer 1974; Korsgaard 2005; Donaldson and Kymlicka 2011). Consequently, relevant animal rights should also be considered when animals are affected by some public policy. Large-scale solar geoengineering would affect animals globally in a myriad of ways, transforming habitats and forcing changes in migratory patterns. If the intervention is terminated too quickly, local extinction risk would increase (Trisos et al. 2018). Of course, increased risk of extinction, degraded habitats and forced migration will also be among the effects of climate change run amok. Thus, the Similar Threat dilemma would arise pertaining to animal rights standing against geoengineering as

well, and the same considerations would apply as in the case of the human rights considered above.

Conferring rights on inanimate natural objects, like mountains, is not a common move, and rights claims made on behalf of nature as such will probably be of marginal significance in the geoengineering discourse. Not even environmental philosophers tend to leverage the "rights" framing. Holmes Rolston III has argued there are "no titles and no laws that can be transgressed in the wilderness", and that rights are conventional (Western) human artefacts, which we could certainly *try* to "project" out of culture onto nature, only to find that the concept would break down, not "attaching" to animals in the wild or other natural objects (Rolston 1988, 48-50). While we disagree on whether rights can successfully be projected onto animals in the wild (I think they can),[53] Rolston's argument is reminiscent of my claim that rights talk is merely shorthand for more basic moral considerations (section 2.1.1). He writes: "There are no rights present in the wild before human assignment. But values (interests, desires, needs satisfied; welfare at stake) may be there apart from the human presence" (ibid., 52). Of course, even if environmental philosophy shies away from the language of rights, someone *could* take as their point of departure the Lovelockian notion of "Gaia" – "the biosphere as an active adaptive control system able to maintain the Earth in homeostasis" (Lovelock and Margulis 1973) – and assert that the Earth has rights as a single, self-regulating entity. Suffice it to say, any such rights claim could only legitimately limit geoengineering if Gaia would suffer more from being engineered than would her occupants otherwise.[54]

5.3.3 Should we engineer the climate despite rights infringement?

The gist of my argument is this: No rights claim is ultimately above the consequences of adhering to it, but rights infringement must never be taken lightly. To sustain the beneficial institution of rights talk, policymakers must – to the best of their ability – seek rigorous justification for any intervention involving rights infringement. Geoengineering

---

[53] At least *negative* and *passive* rights could be applied to all kinds of sentient life (section 2.1). Even if animals are not moral *agents* in the reciprocal sense, they are proper moral patients – even if that may leave us with the massive ethical headache of trying to figure out how we should deal with suffering in the wild (see Horta 2013).

[54] Another suggestion, drawing on the biological philosophy of Hans Jonas (2001), could be that organisms (including plants, algae, amoeba, etc.) have a right to conative expression, i.e. to maintaining their effort to sustain themselves in and from their environments. Of course, no such right – were it to conflict with other rights claims – would be spared from having to justify itself in light of the Utilitarian criterion of rightness.

interventions, like the one imagined in EMERGENCY GEOENGINEERING, could involve rights infringement. However, many of the rights claims assertable against geoengineering might trigger the Similar Threat dilemma, where policymakers would be faced with alternative courses of action all involving similar rights infringements. Here, Bracketed Utilitarianism (on my tentative formulation) prescribes some Rules (of thumb) which *could* tip the scales one way or the other, depending on the scenario. In the less extreme scenarios proposed above, the scale tips in favor of geoengineering. Where no Rule (of thumb) is successful, decision-makers must try to calculate the consequences and choose the alternative which is most likely to yield a significantly better outcome than other available options.

Of course, that geoengineering comes out on top in many of my examples has everything to do with my assumptions about the scenarios. There is still much to be desired in terms of reliable knowledge of possible real-world options and their consequences. Thus, my explorations in this thesis have not yielded positive answers, as much as hinted at patterns of thinking that could be helpful for decision-makers (and perhaps even conducive to the optimal outcome).

I must stress that my EMERGENCY GEOENGINEERING scenario, from which the above argument springs, simply *assumes* that standard mitigation (reducing carbon emissions, and so on) would no longer be a plausible candidate for the expectably optimific solution, due to situational feasibility constraints. At the time when the decision about solar geoengineering is made in my example, it is too late to reduce emissions to a safe level the old-fashioned way. Most likely, this is *not* currently the case (Hausfather and Peters 2020). Thus, if we were considering rights claims standing against geoengineering *today*, the alternative to large-scale solar geoengineering wouldn't be solely runaway climate change, but a range of *other* mitigation strategies as well (in which geoengineering could, or could not, play a role), affecting rights-holders in a plethora of different ways. Assessing all these possible pathways is beyond the scope of this thesis, but Bracketed Utilitarianism could, I think, prove useful for whoever sets out to complete the task. Of course, time *is* running out. We may be facing the EMERGENCY GEOENGINEERING scenario, or something like it, sooner than we'd like. Decision-makers (and moral philosophers) better come prepared.

5.4 Some possible objections

5.4.1 Bad faith

It could be claimed that, even if *rights* don't necessarily limit the justifiability of last-resort geoengineering interventions on their own, other moral considerations help bolster them to that effect. Stephen Gardiner and others have argued that the scenario in which geoengineering becomes necessary, will have been brought about because of "bad faith" and moral corruption in international climate negotiations, and that we must consider the wider ethical context in which geoengineering even becomes a relevant proposal (Gardiner 2013, 12-13; Hamilton 2013, 161-162). Powerful special interest groups, fossil fuel companies and complacent politicians have worked tireless to stall meaningful climate action. Even if this ultimately necessitates desperate measures, wouldn't it be glaringly wrong to infringe on the rights of innocent people in order to clean up the mess? Can we infringe on rights to get out of a bad situation we caused (more or less) knowingly ourselves?

Yes, is the short answer. If we have to. In other words: if the principle of utility clearly demands it. Of course, this does not mean that it wouldn't be regrettable, that obstructionists shouldn't be required to compensate for the damage, or that they shouldn't be publicly condemned. But if the livable biosphere – or the just rights of many *other* innocents – is on the line, we could not presume to forgo necessary action with reference to the wretched nature of history or its antagonists. Regrettable though the measures and their necessity may be, necessary they remain. If, however, we were considering a less dangerous scenario than EMERGENCY GEOENGINEERING, considerations of "bad faith" might count towards choosing other mitigation strategies, ideally not involving rights infringement. Yet, no Consequentialist theory could take the suspect genealogy of some manifest alternative course of action as an overriding moral reason not to pursue it. It would depend on the consequences.

5.4.2 Harmful ways of thinking

It could also be claimed that the kind of thinking underpinning recommendations for geoengineering is, in some ways, harmful. Hamilton (2013, 199), rather polemically, writes: "… how can we think our way out of the problem when the problem is the way we think?". He blames climate change on, among other things, a kind of "technological thinking", which

he associates with "systems analysis, risk assessment and cost-benefit calculation", where the world is understood as "a collection of more or less useable resources" (ibid., 200). This criticism could apply to the kind of rule emphasizing Consequentialist view I have been defending, because deference to net utility demands cost-benefit analysis.

My view is in part inspired by a worry similar to Hamilton's. Utilitarianism is generally put in brackets because Utilitarian thinking could, when badly performed or misused, be harmful (section 2.4). There is, however, no decisive reason to conclude that this type of thinking could not also be helpful, for example with regard to solving our ecological predicament and making other decisions under exceptional circumstances. If we find ourselves in a scenario where (resource-intensive) geoengineering is the only option likely to prevent irreversible damage to the biosphere, we could not reasonably discard this option simply because it would entail viewing the world as, indeed, "useable", nor because our intervention would in the relevant sense be "technological". Had we held such Absolutist views, causing us to dismiss geoengineering out of hand, the biosphere would – in this scenario – perish, undermining what we might have hoped to achieve by promoting other, less "technological", modes of thinking. Of course, Bracketed Utilitarianism might still recommend holistic, nature-deferent modes of thinking as a Rule (of thumb), insofar as this would be expectably optimific.

5.4.3 Barriers to social change

Some people argue that geoengineering and other kinds of "technofix" are – damningly – "conformable with existing structures of power and a society based on continued consumerism." (Hamilton 2013, 175). They typically call for some fundamental social, political and ideological revolution. Hamilton (ibid., 180) writes: "Instead of embarking on a vain quest to mimic the gods, it seems safer and more within our powers to face up to our failures and attempt to become better humans" (Hamilton 2013, 180). He thinks many who gravitate towards technofixes presume, wrongly, that social and political change is inconceivable (ibid., 176). This argument is based on some claims which are, I believe, true:

(A) Many people, especially in positions of power, would like technology to facilitate continued consumerism, because – as they see it – they stand to benefit from this, or because they find alternative modes of organization inconceivable.

(B) Consumerism, insofar as it leads to waste and demands vast amounts of energy, land and resources, is harmful.

(C) Some fundamental changes to our social, political and ideological structures are likely needed.

However, despite the truth of (A), (B) and (C), we would not, on these grounds, have sufficient reason to reject every geoengineering proposal. Although we could reasonably whish that massive social change will occur (or – on some accounts of our predicament – had occurred) before geoengineering becomes (became) a matter of necessity, there are reasons to be pessimistic. Not because massive social change is inconceivable *per se*, but because it may be (or become) inconceivable *at the necessary scale and rate*. The IPCC (2018, 42) notes that, while rapid change has occurred in the past, the geographical and economic scale of the requisite transformation has "no documented historic precedent". Furthermore – as several social theorists have pointed out (e.g., Bourdieu 1972; Giddens 1984) – social structures tend to reproduce. The subtle mechanisms and brute forces opposing revolutionary social, political and ideological transformation should not be underestimated.

Of course, as per the "moral hazard" argument (section 4.3), we may be playing into the hands of these counterrevolutionary forces when we publicly assert the fact that, under some circumstances, geoengineering might be morally permissible. However, as I hope to have showed, this is a fact nonetheless – and one which it may someday prove crucial that policymakers recognize.

6 Concluding remarks

Applying a rule-emphasizing formulation of Act Utilitarianism, which I called Bracketed Utilitarianism, to the question of whether geoengineering might be morally permissible despite rights infringement, I have argued in the affirmative. However, my motivating question was whether there are some rights that may be invoked in opposition to effective climate policies, which ought – as a Rule (of thumb) – to be respected by policymakers in the name of expectably optimific systemic effects. I found that, on some assumptions, plausible candidates for such rights claims would, if invoked against geoengineering, lead to the Similar Threat dilemma, meaning these rights would be threatened either way, whether we engineered the climate or not.

However, the scenarios in which my normative theory yields these results are not yet inevitable. Assuming anthropogenic climate change has not progressed too far for other mitigation strategies to be viable (e.g., rapid, economy-wide decarbonization and fundamental changes to our social, political and ideological structures), Bracketed Utilitarianism could still sanction the invocation of some rights against geoengineering, such as rights to life, liberty, security and self-determination. The infringement of such rights would demand robust justification and should – as a Rule (of thumb) – not be condoned, except if the alternative is some significantly suboptimal outcome (e.g., a sizeable increase in the likelihood of an extinction-level catastrophic event).

Furthermore, Bracketed Utilitarianism would, I believe, not condone the use of geoengineering as a "get out of jail free card" for those eager to sustain unsustainable social structures or lifestyles. To avoid making a problem one is struggling to solve deliberately *worse* seems like a wise Rule (of thumb). As Hamilton (2013, 208-209) puts it:

> The only justification for deploying geoengineering is to make it easier politically to transform our economies and societies so that we live in a way that does not disrupt Earth's natural cycles and the processes that have allowed life to flourish.

If, however, geoengineering should prove necessary to allow this transformation to transpire uninterrupted by the Earth being pushed past some critical climate threshold, causing irreversible damage to the biosphere and human systems, we would need to know as much as possible about what it would and could entail. In keeping with my Rule (of thumb) that policymakers should, if capable, assess every relevant piece of information from reliable sources regarding every possible alternative and outcome before deciding on an emergency intervention, we should condone geoengineering research. Jamieson (2013, 527) said it best:

> Research in areas that involve carbon dioxide removal and solar radiation management should go on as part of the general portfolio of climate-related research, competing with the full panoply of other possible responses to climate change

Hopefully, we need never use geoengineering. I think, however, that we likely will. I count myself among those Soterians who fear that the inertia of social systems – sprinkled as they are with greed and ignorance – will delay the best mitigation strategies for too long. This belief may of course owe itself to my having toiled many years at the frontlines of Norwegian political discourse, advocating for radically different climate policies, without yet seeing my own notions of sufficient action being adopted or operationalized. I may simply be blinded by disillusion and weariness. Perhaps more progress is being made than I am able to appreciate, and things will work out fine. If that's too rosy, perhaps we're at least making sufficient headway that we will avoid manifest climate emergency. If not, then geoengineering it is.

6.1.1 On moral limits

The title of this thesis reads "Moral Limits on Climate Policy". Have I identified some such limit?

Although I have emphasized that Utilitarian calculation should not feature prominently in our decision-making processes, I *did* conjecture that the true theory of normative ethics is a kind of Utilitarianism. If my elaborations on this view are accurate, the overriding moral concern is, ultimately, the sum total of pleasure in the experienceable universe. Theoretically, this view does not play well with moral limits or constraints. It boils everything down to *one* morally relevant normative factor determining the status of a given act, namely its relative contribution to the goodness of outcomes (Kagan 1998, 60). Insofar as our actions are constrained by true morality, it is by the contribution of available options to total goodness.

The rule-emphasizing guise of Bracketed Utilitarianism is justified in light of this normative factor, but Bracketed Utilitarianism allows for no unconditional Rule with a capital R, other than the Rule that the experiences of *every* member of the moral universe count equally toward the goodness of outcomes (cf. Sidgwick 1874, 212).

Hence, there *are* no moral limits constraining climate policy. At least not insofar as the policy in question would *actually* make things better on the whole. This, however, is where Bracketed Utilitarianism kicks in, because humans can never be entirely sure we get it right. Thus, we should impose limits on ourselves, for the sake of the goodness of outcomes, informed by our best guesses of which limits would be optimific. Ideally, these guesses should only be made after rigorous and careful deliberation, attuned to our epistemic shortcomings.

Plausibly, some potential climate policies should be thus constrained. Culling the human population, for example, is impermissible, even if it would reduce carbon emissions, because there's a veritable smorgasbord of viable alternative ways to mitigate global warming, involving significantly more happiness and less suffering. Of course, this is where Deontologists and other commonsense moralists balk at the cold contingency of Utilitarian moral judgements. Isn't culling the population impermissible *full stop*? In the real world, yes. That is probably the best way to look at it. Theoretically? No, regretfully. This much, I think, is true, though it is a bitter pill to swallow. However, given how reluctant policymakers are to even *tax* people in the name of emissions cuts, it seems extremely unlikely that any such barbaric proposal will be made for the sake of climate action. Thus, we need not devote more attention to it.

I am the first to admit that it is a bit scary, facing the possibility of true morality being near-limitless – at least in terms of what kinds of actions it *could* condone, depending on the circumstances. However, I am more afraid of rigid moralities unable to handle exceptional circumstances, and therefore – despite the best of intentions – obstructing possible futures of near-limitless enjoyment (cf. Parfit 2011b, 612-621). My theory, then, tries to accommodate both these worries through the heuristic device of Rules (of thumb) – emphasizing limits, but recognizing that we must sometimes reach beyond them, in order to put the best possible outcome within reach.

References

Anshelm, Jonas, and Anders Hansson. 2016. "Has the grand idea of geoengineering as Plan B run out of steam?" *The Anthropocene Review* 3 (1):64-74. doi: 10.1177/2053019615614592.

Arneson, Richard. 2005. "Sophisticated Rule Consequentialism: Some Simple Objections." *Philosophical Issues* 15 (1):235-251. doi: 10.1111/j.1533-6077.2005.00064.x.

Arrhenius, Gustaf, Jesper Ryberg, and Torbjörn Tännsjö. 2017. "The Repugnant Conclusion." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Stanford University. Original edition, 2006. https://plato.stanford.edu/archives/spr2017/entries/repugnant-conclusion/ (accessed 11.09.2020).

Bentham, Jeremy. 2010. *An Introduction to the Principles of Morals and Legislation (Kindle Edition)*: White Dog Publishing. Original edition, 1781.

Boettcher, Miranda, and Stefan Schäfer. 2017. "Reflecting upon 10 years of geoengineering research: Introduction to the Crutzen + 10 special issue." *Earth's Future* 5:266-277. doi: 10.1002/2016EF000521.

Bostrom, Nick. 2003. "Are We Living in a Computer Simulation?" *The Philosophical Quarterly* 53 (211). doi: 10.1111/1467-9213.00309.

Bostrom, Nick. 2011. "Infinite Ethics." *Analysis and Metaphysics* 10 (2011):9-59.

Bourdieu, Pierre. 1972. *Outline of a Theory of Practice*. Translated by Richard Nice. Edited by Ernest Gellner, Jack Goody, Stephen Gudeman, Michael Herzfeld and Jonathan Parry, *Cambridge Studies in Social and Cultural Anthropology*. Cambridge: Cambridge University Press.

Bradley, Ben. 2006. "Against Satisficing Consequentialism." *Utilitas* 18 (2):97-108. doi: 10.1017/S0953820806001877.

Briggle, Adam. 2018. "Beware of the Toll Keepers: The Ethics of Geoengineering Ethics." *Ethics, Policy & Environment* 21 (2):187-189. doi: 10.1080/21550085.2018.1509479.

Brown, Campbell. 2011. "Consequentialize This." *Ethics* 121 (July 2011):749-771. doi: 10.1086/660696.

Burch-Brown, Joanna M. 2014. "Clues for Consequentialists." *Utilitas* 26 (1):105-119. doi: 10.1017/S0953820813000289.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Climate Action Tracker. 2020. "Temperatures: Addressing global warming." Climate Action Tracker, Last Modified 01.12.2020, accessed 04.12.2020. https://climateactiontracker.org/global/temperatures/.

Cox, Emily, Elspeth Spence, and Nick Pidgeon. 2020. "Public perceptions of carbon dioxide removal in the United States and the United Kingdom." *Nature Climate Change* 10:744-749. doi: 10.1038/s41558-020-0823-z.

Crisp, Roger. 2006. "Hedonism Reconsidered." *Philosophy and Phenomenological Research* 73 (3):619-645. doi: 10.2307/40041013.

Crisp, Roger. 2014. "Taking Stock of Utilitarianism." *Utilitas* 26 (3):231-249. doi: 10.1017/S0953820814000090.

Crisp, Roger, and Morten Kringelbach. 2018. "Higher and Lower Pleasures Revisited: Evidence from Neuroscience." *Neuroethics* 11:211–215. doi: 10.1007/s12152-017-9339-2.

Crutzen, Paul J. 2006. "Albedo Enhancement by Stratospheric Sulfur Injections: A Contribution to Resolve a Policy Dilemma?" *Climatic Change* 77:211-219. doi: 10.1007/s10584-006-9101-y.

Cunsolo, Ashlee, and Neville R. Ellis. 2018. "Ecological grief as a mental health response to climate change-related loss." *Nature Climate Change* 8:275-281. doi: 10.1038/s41558-018-0092-2.

Dennett, Daniel C. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Company, Inc.

Descartes, René. 2017. "Meditations on First Philosophy: Meditation I." In *Collected Works of René Descartes (Kindle ed.)*. Delphi Classics. Original edition, 1641.

Donaldson, Sue, and Will Kymlicka. 2011. *Zoopolis : a political theory of animal rights*. Oxford: Oxford University Press.

Dreier, James. 1993. "Structures of Normative Theories." *The Monist* 76 (1):22-40. doi: 10.5840/monist19937616.

Ellis, Erle C. 2018. *Anthropocene: A Very Short Introduction*, *Very Short Introductions*. Oxford: Oxford University Press.

Fjellberg, Anders. 2018. "Brent Jord III." *Dagbladet*, Klima. Accessed 10.10.2018. https://www.dagbladet.no/nyheter/jeg-tror-ikke-mange-mener-det-er-mulig-sier-klimaforsker-glen-peters/70285139.

Fox, Andrew S., Regina C. Lapate, Alexander J. Shackman, and Richard J. Davidson, eds. 2018. *The Nature of Emotion: Fundamental Questions*. 2nd ed, *Series In Affective Science*. New York: Oxford University Press.

Frederick, Danny. 2014. "Pro-tanto Versus Absolute Rights." *Philosophical Forum* 45 (4):375-394. doi: 10.1111/phil.12044.

Freeman, Samuel. 2016. Original Position. In *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*, edited by Edward N. Zalta.

Gardiner, Sephen M. 2013. "Geoengineering and Moral Schizophrenia: What Is the Question?" In *Climate Change Geoengineering*, edited by Wil C. G. Burns and Andrew L. Strauss, 11-38. Cambridge: Cambridge University Press.

Gasper, Karen, Lauren A. Spencer, and Danfei Hu. 2019. "Does Neutral Affect Exist?: How Challenging Three Beliefs About Neutral Affect Can Advance Affective Research." *Frontiers in Psychology* 10 (2476):1-11. doi: 10.3389/fpsyg.2019.02476.

Giddens, Anthony. 1984. *The Constitution of Society. Outline of the Theory of Structuration*: Polity Press.

Goodell, Jeff. 2010. *How to Cool the Planet: Geoengineering and the Audacious Quest to Fix Earth's Climate*. Boston/New York: Houghton Mifflin Harcourt.

Hamilton, Clive. 2013. *Earthmasters: The Dawn of the Age of Climate Engineering* New Haven and London: Yale University Press.

Hare, R. M. 1981. *Moral Thinking: Its Levels, Method and Point*. Oxford: Clarendon Press.

Hare, R. M. 1997. *Sorting out Ethics*. Oxford: Clarendon Press.

Harel, Alon, and Assaf Sharon. 2011. "'Necessity knows no law': On Extreme Cases and Uncodifiable Necessities." *University of Toronto Law Journal* 4:845-865. doi: 10.1353/tlj.2011.0037.

Hausfather, Zeke, and Glen P. Peters. 2020. "Emissions – the 'business as usual' story is misleading." *Nature* 577 (7792). doi: 10.1038/d41586-020-00177-3.

Hobbes, Thomas. 1996. *Leviathan*. Edited by J. C. A. Gaskin, *Oxford World's Classics*. Oxford: Oxford University Press.

Hoffman, Donald D. 2019. "Did we evolve to see reality, or are spacetime and objects just our user interface?" *Annals of the New York Academy of Sciences: Special Issue: Conversations on the Nature of Reality* 1458 (2019):65-69. doi: 10.1111/nyas.14222.

Hohfeld, Wesley Newcomb. 1917. "Fundamental Legal Conceptions as Applied in Judicial Reasoning." *The Yale Law Journal* 26 (8):710-770. doi: 10.2307/786270.

Horta, Oscar. 2013. "Zoopolis, Interventions and the State of Nature." *Law, Ethics and Philosophy* (1):113-125.

Huemer, Michael. 2008. "In Defence of Repugnance." *Mind* 117 (468):899-933. doi: 10.1093/mind/fzn079.

IPCC. 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by Core Writing Team, R.K. Pachauri and L.A. Meyer. Geneva, Switzerland: IPCC.

IPCC. 2018. *Global warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of*

climate change, sustainable development, and efforts to eradicate poverty. Edited by Masson-Delmotte, P. Zhai V., H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor and T. Waterfield. In press.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32 (127):127-136.

Jackson, Frank. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101 (3):461-482. doi: 10.1086/293312.

Jamieson, Dale. 2007. "When Utilitarians Should Be Virtue Theorists." *Utilitas* 19 (2):160-183. doi: 10.1017/S0953820807002452.

Jamieson, Dale. 2013. "Some whats, whys and worries of geoengineering." *Climatic Change* 121:527-537. doi: 10.1007/s10584-013-0862-9.

Jonas, Hans. 2001. *The Phenomenon of Life*. Illinois: Northwestern University Press. Original edition, 1966.

Jones, Nicola. 2017. "How the World Passed a Carbon Threshold and Why It Matters." *Yale Environment 360*. Accessed 03.12.2020. https://e360.yale.edu/features/how-the-world-passed-a-carbon-threshold-400ppm-and-why-it-matters

Kagan, Shelly. 1998. *Normative Ethics*. Edited by Norman Daniels and Keith Lehrer, *Dimensions of Philosophy Series*. Colorado: Westview Press.

Keith, David. 2013. *A Case for Climate Engineering*, *Boston Review*: MIT Press.

Kirk, Robert. 2019. "Zombies." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/zombies/ (accessed 11.04.19).

Korsgaard, Christine M. 2005. "Fellow Creatures: Kantian Ethics and Our Duties to Animals." *Tanner Lectures on Human Values* 25:77-110.

Kraut, Richard. 2018. "Aristotle's Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/ (accessed 10.12.2020).

Lazari-Radek, Katarzyna de, and Peter Singer. 2010. "Secrecy in Consequentialism: A Defence of Esoteric Morality." *Ratio* 23 (1):34-58. doi: 10.1111/j.1467-9329.2009.00449.x.

Lazari-Radek, Katarzyna de, and Peter Singer. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.

Lazari-Radek, Katarzyna de, and Peter Singer. 2017. *Utilitarianism: a very short introduction*, *Very short introductions*. Oxford: Oxford University Press.

Lenman, James. 2000. "Consequentialism and CIuelessness." *Philosophy & Public Affairs* 29 (4):342-370.

Lenton, Timothy M., Hermann Held, Elmar Kriegler, Jim W. Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. 2008. "Tipping elements in the Earth's climate system." *PNAS* 105 (6):1786-1793. doi: 10.1073/pnas.0705414105.

Lenton, Timothy M., Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen, and Hans Joachim Schellnhuber. 2019. "Climate tipping points — too risky to bet against." *Nature* 575 (2019):592-596. doi: 10.1038/d41586-019-03595-0.

Lindsey, Rebecca. 2020. "Climate Change: Atmospheric Carbon Dioxide." NOAA, Last Modified 14.08.2020, accessed 03.12.2020. https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide.

Louise, Jennie. 2006. "Right Motive, Wrong Action: Direct Consequentialism and Evaluative Conflict." *Ethical Theory and Moral Practice* 9 (1):65-85. doi: 10.1007/sl0677-005-9000-8.

Lovelock, James E., and Lynn Margulis. 1973. "Atmospheric homeostasis by and for the biosphere: the gaia hypothesis." *Tellus* 26 (1-2):2-10. doi: 10.1111/j.2153-3490.1974.tb01946.x.

Lynas, Mark. 2020. *Our Final Warning: Six Degrees of Climate Emergency*. London: 4th Estate.

Mark, Jason. 2018. "The Case for Climate Reparations: Who should pay the costs for climate-change-related disasters?". Sierra Club, Last Modified 23.04.2018, accessed 07.12.2020. https://www.sierraclub.org/sierra/2018-3-may-june/feature/the-case-for-climate-reparations.

Mauritsen, Thorsten, and Robert Pincus. 2017. "Committed warming inferred from observations." *Nature Climate Change* 7 (2017):652-655. doi: 10.1038/nclimate3357.

Maynard, Jonathan Leader. 2014. "Rethinking the Role of Ideology in Mass Atrocities." *Terrorism and Political Violence* 26 (5):821-841. doi: 10.1080/09546553.2013.796934.

Moore, George Edward. 1922. "Principa Ethica." London: Cambridge University Press. https://www.gutenberg.org/files/53430/53430-h/53430-h.htm (accessed 02.09.2020).

Moore, John C., Ying Chen, Xuefeng Cui, Wenping Yuan, Wenjie Dong, Yun Gao, and Peijun Shi. 2016. "Will China be the first to initiate climate engineering?" *Earth's Future* 4:588-595. doi: 10.1002/2016EF000402.

Murphy, Thérèse, and Noel Whitty. 2009. "Is Human Rights Prepared? Risk, Rights and Public Health Emergencies." *Medical Law Review* 17 (Summer 2009):219-244. doi: 10.1093/medlaw/fwp007.

Nagel, Jennifer. 2014. *Knowledge: A Very Short Introduction*, *A Very Short Introduction*. Oxford: Oxford University Press.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83, (4): 435-450.

Nagel, Thomas. 1979a. "Subjective and Objective." In *Mortal Questions*, 196-215. Cambridge: Cambridge University Press.

Nagel, Thomas. 1979b. "War and Massacre." In *Mortal Questions*, 53-74. Cambridge: Cambridge University Press.

Nalam, Aditya, Govindasamy Bala, and Angshuman Modak. 2018. "Effects of Arctic geoengineering on precipitation in the tropical monsoon regions." *Climate dynamics* 50 (9):3375-3395. doi: 10.1007/s00382-017-3810-y.

NASA. 2020. "The Causes of Climate Change." Earth Science Communications Team at NASA's Jet Propulsion Laboratory, Last Modified 29.10.20, accessed 02.11.20. https://climate.nasa.gov/causes/.

Norgaard, Kari Marie. 2011. *Living in Denial: Climate Change, Emotions and Everyday Life*. Cambridge, MA: The MIT Press.

Norman, Emma S. 2017. "Standing Up for Inherent Rights: The Role of Indigenous-Led Activism in Protecting Sacred Waters and Ways of Life." *Society & Natural Resources* 30 (4):537-553. doi: 10.1080/08941920.2016.1274459.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell.

NOAA. 2020. "Trends in Atmospheric Carbon Dioxide." Global Monitoring Laboratory: Earth System Research Laboratories, Last Modified 06.11.2020, accessed 03.12.2020. https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html#global.

Oberdiek, John. 2008. "Specifying Rights Out of Necessity." *Oxford Journal of Legal Studies* 28:127-146. doi: 10.1093/ojls/gqm028.

Olivier, J.G.J., and J.A.H.W. Peters. 2020. Trends in global CO2 and total greenhouse gas emissions: 2019 Report. The Hague: PBL Netherlands Environmental Assessment Agency.

Pappas, Stephanie. 2017. "Facts About Sulfur." Last Modified 29.09.2017, accessed 04.12.2020. https://www.livescience.com/28939-sulfur.html.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, Derek. 2011a. *On What Matters: Volume One*. Edited by Samuel Scheffler. 3 vols. Vol. 1, *The Berkeley Tanner Lectures*. Oxford: Oxford University Press.

Parfit, Derek. 2011b. *On What Matters: Volume Two*. Edited by Samuel Scheffler. 3 vols. Vol. 2, *The Berkeley Tanner Lectures*. Oxford: Oxford University Press.

Parfit, Derek. 2017. *On What Matters: Volume Three*. 3 vols. Vol. 3. Oxford: Oxford University Press.

Parson, Edward A. 2014. "Climate Engineering in Global Climate Governance: Implications for Participation and Linkage." *Transnational Environmental Law* 3 (1):89-110. doi: 10.101 7/2047102513000496.

Persson, Ingmar. 2008. "A Consequentialist Distinction between What We Ought to Do and Ought to Try." *Utilitas* 20 (3):348-355. doi: 10.1017/S0953820808003191.

Portmore, Douglas. 2007. "Consequentializing Moral Theories." *Pacific Philosophical Quarterly* 88 (1):39-73. doi: 10.1111/j.1468-0114.2007.00280.x.

Portmore, Douglas. 2009. "Maximizing and Satisficing Consequentialism." PhilPapers.org, Last Modified 29.06.2020, accessed 25.08.2020. https://philpapers.org/browse/maximizing-and-satisficing-consequentialism.

Pressman, Michael. 2015. "A Defence of Average Utilitarianism." *Utilitas* 27 (4):389-424. doi: 10.1017/S0953820815000072.

Prinz, Jesse. 2007. *The Emotionial Construction of Morals*. New York: Oxford University Press.

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 13 (2): 134-171. Accessed 14.08.2020. www.jstor.org/stable/2265273

Rasch, Philip J, John Latham, and Chih-Chieh (Jack) Chen. 2009. "Geoengineering by cloud seeding: influence on sea ice and climate system." *Environmental Research Letters* 4 (045112):1-8. doi: 10.1088/1748-9326/4/4/045112.

Rawls, John. 1999. *A Theory of Justice*. Rev. ed. Cambridge: The Belknap Press of Harvard University Press. Original edition, 1971.

Reynolds, Jesse L., Andy Parker, and Peter Irvine. 2016. "Five solar geoengineering tropes that have outstayed their welcome." *Earth's Future* 4:562-568. doi: 10.1002/2016EF000416.

Risse, Mathias. 2009. "The Right to Relocation: Disappearing Island Nations and Common Ownership of the Earth." *Ethics & International Affairs* 23 (3):281-300.

Robock, Alan. 2008. "20 reasons why geoengineering may be a bad idea." *Bulletin of the Atom Scientists* 64 (2):14-18, 59. doi: 10.2968/064002006.

Rogelj, J. , D. Shindell, K. Jiang, S. Fifita, P. Forster, V. Ginzburg, C. Handa, H. Kheshgi, S. Kobayashi, E. Kriegler, L. Mundaca, R. Séférian, and M.V.Vilariño. 2018. "Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development." In *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, edited by Masson-Delmotte, P. Zhai V., H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, X. Zhou Y. Chen, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor and T. Waterfield, 93-174. In Press.

Rolston, Holmes. 1988. *Environmental Ethics: Duties to and Values in The Natural World*. Edited by Tom Regan, *Ethics and Action*. Philadelphia: Temple University Press.

Rowell, David, and Luke B. Connelly. 2012. "A History of the Term «Moral Hazard»." *The Journal of Risk and Insurance* 79 (4):1051-1075. doi: 10.1111/j.1539-6975.2011.01448.x.

Roy, Shouraseni Sen. 2018. "Climate Change in the Global South: Trends and Spatial Patterns." In *Linking Gender to Climate Change Impacts in the Global South*, 1-26. Springer International Publishing. doi:10.1007/978-3-319-75777-3.

Royal Society. 2009. "Geoengineering the climate: science, governance and uncertainty." In *RS Policy Document 10/09*. London: The Royal Society. https://royalsociety.org/topics-policy/publications/2009/geoengineering-climate/ (accessed 05.12.2020).

Ryberg, Jesper. 1996. "Is the Repugnant Conclusion Repugnant?" *Philosophical Papers* 25 (3):161-177. doi: 10.1080/05568649609506547.

Sandin, Per, and Misse Wester. 2009. "The Moral Black Hole." *Ethical Theory and Moral Practice* 12:291-301.

Sidgwick, Henry. 1874. *The Methods of Ethics*. London: Macmillan and co.

Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1 (3):229-243.

Singer, Peter. 1974. "All Animals Are Equal." *Philosophic Exchange* 5 (1):103-116.

Singer, Peter. 2005. "Ethics and intuitions." *The Journal of Ethics* 9:331-352.

Singer, Peter. 2019. *The Life You Can Save: How To Do Your Part To End World Poverty (10th Anniversary Edition)*. Edited by The Life You Can Save. Kindle ed. Washington: The Life You Can Save. Original edition, 2009.

Sinnott-Armstrong, Walter. 2015. "Consequentialism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Stanford University. Original edition, 2003. http://plato.stanford.edu/archives/win2015/entries/consequentialism/ (accessed 29.09.18).

Smart, J.J.C. 1973. "An Outline of a System of Utilitarian Ethics." In *Utilitarianism: For and Against*. New York: Cambridge University Press.

Sorell, Tom. 2003. "Morality and Emergency." *Proceedings of the Aristotelian Society* 103:21-37.

Sterri, Aksel Braanen, and Ole Martin Moen. 2020. "The Ethics of Emergencies." *Philosophical studies*. doi: 10.1007/s11098-020-01566-0.

Tong, Dan, Qiang Zhang, Yixuan Zheng, Ken Caldeira, Christine Shearer, Chaopeng Hong, Yue Qin, and Steven J. Davis. 2019. "Committed emissions from existing energy

infrastructure jeopardize 1.5 °C climate target." *Nature* 572 (2019):373-377. doi: 10.1038/s41586-019-1364-3.

Trisos, Christopher H., Giuseppe Amatulli, Jessica Gurevitch, Alan Robock, Lili Xia, and Brian Zambri. 2018. "Potentially dangerous consequences for biodiversity of solar geoengineering implementation and termination." *Nature Ecology & Evolution* 2:475-482. doi: 10.1038/s41559-017-0431-0.

UN General Assembly. 1948. "The Universal Declaration of Human Rights." United Nations, accessed 10.12.2020. https://www.un.org/en/universal-declaration-human-rights/index.html.

UNEP. 2020. *Emissions Gap Report 2020*. Nairobi: United Nations Environment Programme.

Wagner, Gernot, and Martin L. Weitzman. 2015. *Climate Shock: The Economic Consequences of a Hotter Planet*. Oxford: Princeton University Press.

Weinberg, Jonathan M. 2016. "Intuitions." In *The Oxford Handbook of Philosophical Methodology*, edited by Herman Cappelen, Tamar S. Gendler og John Hawthorne, 287-308. Oxford: Oxford University Press.

Wenar, Leif. 2015. "Rights." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. http://plato.stanford.edu/archives/fall2015/entries/rights/ (accessed 01.10.19).

Williams, Bernard. 1973. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*. New York: Cambridge University Press.

Williams, Bernard. 1995. "The point of view of the universe: Sidgwick and the ambitions of ethics." In *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*, 153-171. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621246.

Wunderling, Nico, Matteo Willeit, Jonathan F. Donges, and Ricarda Winkelmann. 2020. "Global warming due to loss of large ice masses and Arctic summer sea ice." *Nature Communications* 11 (5177). doi: 10.1038/s41467-020-18934-3.

Ypi, Lea. 2012. "A Permissive Theory of Territorial Rights." *European Journal of Philosophy* 22 (2):288-312. doi: 10.1111/j.1468-0378.2011.00506.x.