

PARFITIAN CONVERGENCE AND THE STRUCTURE OF MORAL THEORIES

by

Berker Bahçeci

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for

the Degree of

Master of Arts

in

Philosophy

Koç University

3 September 2024

PARFITIAN CONVERGENCE AND THE STRUCTURE OF MORAL THEORIES

Koç University

Graduate School of Social Sciences and

This is to certify that I have examined this copy of a master's thesis by

Berker Bahçeci

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Asst. Prof. David Killoren (Advisor)

Prof. Zeynep Direk

Dr. Marius Baumann

Date: 24.09.2024

ABSTRACT

Derek Parfit argues competing moral traditions of Kantianism, consequentialism, and contractualism, in their best formulations, converge to one another. In this work, I focus on clarifying what this convergence thesis means. As Parfit is elusive at points regarding convergence, I interpret his mountain metaphor to extract plausible insights, and analyze the convergence thesis. I put forward a structure of moral theories, which I argue blocks total convergence due to unresolvable explanatory differences. I put forward a weaker version of the convergence thesis which leads to the conclusion that Parfit nonetheless has achieved much in resolving the disagreement among moral experts. I conclude by responding an argument due to Marius Baumann, and argue that unlike Parfit's convergence thesis, the weaker convergence thesis I put forward is not threatened by underdetermination of moral theories by evidence.

ÖZETÇE

Derek Parfit Kantçılığın, sonuççuluğun, ve kontratçılığın en iyi formülasyonlarının birbirlerine yakınsadıklarını savunuyor. Bu çalışmada, bu yakınsamanın ne anlama geldiğini aydınlatmak üzerine yoğunlaşıyorum. Parfit yakınsama hususunda zaman zaman kaçamaklı olduğundan, yakınsamaya dair olası vukufiyetlere varmak ve yakınsama tezini analiz etmek için Parfit'in dağ metaforunu yorumluyorum. Etik teorilere dair bir teorik yapı öne sürüyorum, ve açıklamasal farklılıkların çözülemeyecek olduğunu savunarak tekmil yakınsama tezini reddediyorum. Yakınsamaya dair daha zayıf bir tez öne sürüyorum, ve bu zayıf tez ışığında Parfit'in etik uzmanlar arasındaki anlaşmazlığı çözme yolunda yine de uzun bir yol kat etmiş bulunduğunu savunuyorum. Çalışmamı Marius Baumann tarafından öne sürülen etik teorilerin delil tarafından az-belirlenmesi argümanının öne sürdüğüm zayıf yakınsama tezine bir cevap olarak öne sürerek tamamliyorum.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, David, who took me and this project on-board on short notice. He provided more support and guidance than I could have imagined, for which I am thankful. I would also like to express my gratitude to my partner, Çağla, whose support, companionship, kindness, and thoughtfulness helps me go through every challenge I face.

This project is partially funded by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) 2210/B scholarship program. I would like to express my gratitude to TÜBİTAK for their support.

TEŞEKKÜRLER

Beni ve bu proje üstünde çalışmamı kısa zamanda kabul eden danışmanım David'e teşekkürleri sunmak isterim. Bana hayal edebileceğimden çok daha fazla destek olup yol gösterdiği için kendisine müteşekkirim. Partnerim Çağla'ya da desteği, yoldaşlığı, kibarlığı, ve düşünceliliğiyle karşıma çıkan her engeli aşmamda yardımcı olduğu için minnettarım.

Bu proje Türkiye Bilimsel ve Teknolojik Araştırma Kurumu'nun (TÜBİTAK) 2210/B yüksek lisans bursu tarafından desteklenmiştir. TÜBİTAK'a destekleri için teşekkürlerimi sunarım.

TABLE OF CONTENTS

Chapter 1: Introduction	p.8
Chapter 2: Parfit's Project	p.11
Chapter 3: The Convergence Claim	p.17
3.1 The Mountain Metaphor	p.17
3.2 Convergence in Scientific Theorizing	p.20
3.3 The Structure of Moral Theories	p.34
Chapter 4: The Narrow Convergence Claim and TT	p.45
Chapter 5: Underdetermination of Moral Theories	p.55
5.1 Moral Evidence	p.57
5.2 Underdetermination of TT and NCC''	p.60
Chapter 6: Conclusion	p.64
References	p.66

Chapter 1: Introduction

In *On What Matters*, Derek Parfit (2011) argues that three main traditions of normative ethics, upon reflection and modification, can be “combined” (2011, p.412) into a single theory. He calls this theory *The Triple Theory* (2011, p.412). Historically, these three traditions have been seen as competitors which are logically inconsistent with one another. According to Kantianism, an act is right if it is allowed by universally willable principles. Contractualists hold that an act is right if it is allowed by principles that no one can reasonably reject. Consequentialists argue that an act is right if it is allowed by optimific principles. Despite the tension among these traditions, even in their simplest formulation, there has been efforts to reconcile, or bring these traditions closer. Hare (1993) for instance argues Kantianism and utilitarianism, a strain of consequentialism, can be brought together. Scanlon (1982) considers whether contractualism and utilitarianism can be reconciled, whereas Portmore (2023) engages in a similar effort with respect to Kantianism and consequentialism.

Parfit goes a step further in his effort in that compared to the three above examples, he tries to bring together three competing traditions rather than two. He claims these traditions are, in his evocative metaphor, “climbing the same mountain on different sides” (2011, p.419). Call this the *Convergence Claim*. Unfortunately, the argument for the Triple Theory (henceforth TT) and the Convergence Claim (henceforth CC) takes place rather quickly, and the exact nature of TT is not sufficiently clear. Is TT a distinct theory from, say, contractualism; and if so, how does it distinguish itself from contractualism? What exactly converges to what, and how? Is TT a genuine moral theory? What is the difference between CC and TT? Can CC be true without TT being true? These are questions waiting for an answer. What we do know is that Parfit thinks a handful of

moral principles “describe a single, high-order wrong-making property, under which all other such properties can be subsumed” (2011, p.414) which TT describes, and that the same set of moral principles can be found in Parfit’s preferred versions of Kantianism, consequentialism, and contractualism.

Parfit’s project, like any philosophical project, has received criticism.¹ Much of these criticisms are directed at Parfit’s construction of particular theories; that is, they criticize Parfit on grounds that his construction of, say, his preferred version of Kantianism is not Kantian at heart. Rather, critics claim, it is a Kantianism formulated to be consequentialized without proper Kantian commitments (Wolf, 2011). Another critic takes a different approach. Baumann (2018) assumes that CC obtains, without committing to whether CC *actually* obtains, and argues that the choice of moral theory is underdetermined by evidence. If the competing theories yield the same *deontic verdicts*, then there is no way for us to tell which one is true based on the evidence available to us.

CC raises interesting questions regarding the nature and structure of moral theories, and an exposition of Parfit’s elusive mountain metaphor is an opportunity to dwell upon the structure of moral theories. Baumann’s criticism, on the other hand, invites us to think about the moral epistemology relevant for assessing moral theories. The central aim of this thesis is to investigate the nature and epistemology of moral theories, with a focus on Parfit. To do so, I will take an unconventional path. Building on the work of Dietrich and List (2017), and Baumann (2018) who

¹ For instance, objections of Herman (2011) and Scanlon (2011) revolve around their views not being properly presented. They claim that if their views are properly represented, TT would not have obtained. Chappell (2021) is skeptical of whether convergence is desirable for moral theorizing, or that it is indicative of progress. Wolf (2011) argues aims of moral theorizing go beyond converging on some substantial moral truths, and as such, CC is reflective of a mistaken meta-ethical view.

uses distinctions from philosophy of science, I will employ distinctions about the structure of theories, evidence and confirmation that have been thus far widely discussed in philosophy of science and apply them to the moral domain.

The structure of this work will be as follows: In section 2, I will give an overview of Parfit's project. In section 3, I will expand upon what exactly *convergence* means. I will argue that Parfit aims for a what I'll call a *narrow* convergence of competing moral theories, which is, as I'll explain in more detail, a less demanding convergence claim with a narrower region of agreement. If narrow convergence is true, then significant differences remain between Kantianism, consequentialism, and contractualism, and TT. In section 4, I will highlight the differences between narrow convergence and TT, and argue that while the former obtains, the latter does not. In section 5, I will present a response to Baumann (2018). I will first clarify the moral evidence in operation in Baumann's work, and then argue that although TT is underdetermined, narrow convergence is not; for narrow convergence allows for differences between competing theories which can break the tie when it comes to deciding which one of the theories is true.

Chapter 2: Parfit's Project

In understanding Parfit's claim for convergence, it is beneficial to understand what Parfit's aim is in arguing for CC. Why is Parfit concerned to show that different moral theories *converge*? Parfit is troubled by the widespread disagreement on ethics. What seems to bother him even more is that the disagreement is among experts, which cannot be explained away by appealing to biases, interests, background beliefs, and the like (Parfit, 2011, p.418; 2011b, p.428). The argument that worries him can be structured as:

- D1. If there are some moral facts that moral theorizing is about, then there should be considerable agreement among experts.
- D2. There is not considerable agreement among experts.
- D3. Therefore, there are no moral facts that moral theorizing is about.

If there are some moral facts that moral theorizing is about, which Parfit believes there are, then one would expect the experts (e.g., moral philosophers) to reach at least some agreement. Who are moral experts? Alas, Parfit does not give us an answer; the word does not appear in *On What Matters*. The epistemological literature on arguments from disagreement often revolves around idealized epistemic peers, thus by-passing the possible issues relating to expertise (McGrath, 2008, p.97). Furthermore, as McGrath (2008, p.97) points out, it is difficult to ascribe moral expertise to someone without having some independent criteria by which we can tell who is getting things right. Without knowing what the morally right thing to do is in most cases, and without being viciously circular, there is no obvious way to assign moral expertise to any group of people (McGrath 2008, p.98). However, for the present purposes, we know that Parfit's locutors are his fellow moral philosophers. He is not so much concerned about the moral disagreement among, say, plumbers,

athletes, students, etc. So, within the scope of this work, let us stipulate that moral experts are “individual moral theorists as exponents of distinctively Kantian, contractualist, and consequentialist traditions” (Wolf, 2011, p.51). I believe nothing would change if we were to simply replace moral experts with moral peers, other than to further argue that the disagreement cannot be resolved by appealing to a different set of peers who are more knowledgeable, or more skilled, or less biased in the area.

Moral philosophers, conceived as experts, are “(...) familiar with moral concepts and with moral arguments, who has ample time to gather information and think about it, may reasonably be expected to reach a soundly based conclusion.” (Singer and Wells, 1984, p.177) Non-experts might lack some of the skills, like conceptual understanding and bias avoidance, and knowledge the moral expert has, which might cause them to err in their reasoning, or talk past each other. Plausibly, different moral philosophers using the same concepts, arguments, and information will, to a certain extent, reach the same conclusions more frequently compared to the non-experts; hence the expectation that experts should reach at least some agreement. (Bina, Bonicalzi & Croce, forthcoming). Yet, our most promising ethical theories, endorsed by different experts, tell us different stories when it comes to what we ought to do, or so it seems. So, the argument from disagreement goes, there are no moral facts that moral theorizing is about.² In reading and interpreting Parfit, we must keep in mind that his aim is to resolve at least some of the disagreement,

² It is worth noting that this claim does not entail that there are no facts of the matter about morality. It is possible that there are moral truths which our best theories simply cannot capture. An analogous example would be Aristotelean theory of gravity, according to which objects fall because they tend towards their ‘natural state’. Today, we know that the fact of the matter is that bodies interact with gravitational fields, which the Aristotelean theory completely misses. So, Parfit’s worry is not that there are no moral facts, but rather that our best efforts to systematically understand and explain such truths is in vain.

to argue that there are at least some moral truths on which the experts can agree and that the region of agreement is larger than many seem to think.

Parfit's strategy to justify his claim that there are shared moral truths across different moral theories is to show that there are shared moral truths across different moral theories; hence removing or reducing the disagreement among experts. To do so, he picks what he considers to be best and most insightful versions of Kantianism, contractualism, and consequentialism, and argues that these distinct theories are crawling towards a "supreme principle of morality" (Wolf, 2011, p.34). By reflection and modification, reconciling the tensions between the competing theories, Parfit (2011, p.419) claims, allows us to discover that these theories are "climbing the same mountain on different sides."

His effort begins in chapter 15 of *On What Matters*. After criticizing Rawls's formula for contractualism, he identifies two contractualist formulas which allow Parfit to "reach a more plausible and successful version of Contractualism." (p.355) They are:

The Kantian Contractualist Formula: Everyone ought to follow the principles whose universal acceptance everyone could rationally will, or choose. (p.355)

Scanlon's Formula: An act is wrong just when such acts are disallowed by some principle that no one could reasonably reject. (p.369)

Then, Parfit turns his attention to consequentialism, and initially formulates Rule Consequentialism as:

RC: Everyone ought to follow the principles whose universal acceptance would make things go best. (p.375)

Through chapter 16 and the beginning of chapter 17, he argues that RC triumphs over other plausible formulations of consequentialism, and through modifications, arrives at his preferred version of Rule Consequentialism formulated as:

RC': Everyone ought to follow (these) optimific principles.³ (p.410)

RC' then states, according to Parfit, that we should follow principles which would, all things considered, yield the best outcome, which also are the same principles that everyone can rationally will to be universal laws. Parfit argues RC' is implied by The Kantian Contractualist Formula, and proposes to combine them into

Kantian Rule Consequentialism: Everyone ought to follow the optimific principles, because these are the only principles that everyone could rationally will to be universal laws. (p.411)

The penultimate step in his argument is to argue that The Kantian Contractualist Formula can be combined with Scanlon's Contractualist Formula, which centers around the notion of not being reasonable rejectable by someone. If this is true, then Parfit claims that "the principles that no one could reasonably reject are the same as the principles that everyone could rationally will to be universal laws." (p. 412) To do so, Parfit claims that if we can't rationally will a principle to be a universal law, then there must be a strong objection to this principle. Since Kantian Contractualist Formula implies that we can rationally will some principles, then there cannot be a strong objection to these principles, that these principles are not reasonably rejectable on the basis of a strong objection.

³ The parenthetical (these) is referring to principles "that everyone could rationally will to be universal laws" (p.410)

Since the Kantian Contractualist Formula, which is now combined with Scanlon's formulation, also implies Rule Consequentialism, Parfit combines them all into a single theory; namely, the *Triple Theory*:

TT: An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable. (p.413)

Principles of TT are, according to Parfit "triply supported" (p.413), meaning that such principles are found in Parfit's preferred versions of Kantianism, contractualism and consequentialism. Crucially, Parfit claims that all principles of these theories are all triply supported, that "these three properties (*optimificity, uniquely universally willableness, and not being reasonable rejectable*) are all had by all and only the same principles." (p.413) In the face of a plausible question of comparing, say, the moral importance of optimificity and not being reasonably rejectable, Parfit claims that such a question would not arise. If Parfit's claim is true, then any *prima facie* optimific principle that captures at a general level some moral truth is at the same time uniquely universally willable, and not reasonably rejectable.⁴ Likewise, any *prima facie* not reasonably rejectable principle is uniquely willable and optimific; and the same permutation holds for *prima facie* uniquely universally willable principles.

However, the success of Parfit's project is assessed, the effort is surely monumental. If Parfit is right, then Kantianism, contractualism, and consequentialism all get to the same moral truths, *in some*

⁴ Here, I use *prima facie* in such a sense that at a first look at some principle P, P strikes us, or seems to us, as optimific. Upon reflection, we might come to the conclusion that P is not merely optimific, but also universally willable, and not reasonably rejectable, which, I believe, what Parfit claims to be the case.

way. If that's the case, then the experts which endorse these different theories would, using the theories themselves, frequently reach the same conclusions. Frequently reaching the same conclusions would be sufficient to reject the claim that there is no considerable agreement among the experts, namely D2. Hence, the argument from disagreement is dissolved or the very least, loses much of its force, and moral theorizing is no longer under its threat.

The final argumentative maneuver that gets Parfit where he wants, i.e., to show that there is considerable agreement among experts thanks to convergence, however, is rather quick, and leaves a lot unexplained. Parfit's conception of a moral theory seems to primarily revolve around the principles a theory endorses, but we can ask whether that is all, or primarily is, what defines a moral theory. Perhaps, the more important question is in what way exactly these theories get to the same truth, and which moral truths do they get to. Answering these questions will allow us to better assess Parfit's success. In the next two sections, we will answer these questions by taking a closer look at the elusive mountain metaphor, and the structure of moral theories.

Chapter 3: The Convergence Claim

The claim that Parfit's preferred versions of Kantianism, contractualism, and consequentialism *converge* (i.e., CC) is distinct from the claim that TT is true, or that they converge into TT. It is possible that these theories converge into something else, or that they converge in a particular way which does not entail TT. CC might be true, and TT might be false. Assessing whether Parfit succeeds in his aim in *On What Matters* partially relies on having a good grasp of what is the final product. Although it is easy to think that this end product is that the three theories Parfit handles converge into TT, and assess CC and TT to be on par, this need not be the case. I believe CC is a much more plausible claim than TT, and that Parfit's monumental effort should partially be assessed based on whether CC is true or not, rather than TT. Furthermore, some form of CC can be sufficient to defeat the argument from disagreement. In this section and the next, I will dwell upon CC and TT respectively, and expand upon them. I will argue that in light of Parfit's aim and metaphor, CC obtains whereas TT falls short as a genuine moral theory.

3.1: The Mountain Metaphor

Let us begin by recalling the metaphor. Three rival moral theories are climbing the same mountain, on different sides. The metaphor has three components about which we can have questions: the mountain, its sides, and the act of climbing. What is the significance of the different sides? What does climbing symbolize? What does it mean for an act of climbing to be a success? How does this all relate to moral theories, and how does this help us better understand CC?

In the literature on *On What Matters*, little interpretive work has been done on the mountain metaphor. As such, it is not easy for me to claim, other than making my case for it, that my

interpretation is a plausible one, which can be further justified by appealing to, say, similar interpretations. One author who engages with the metaphor is Wolf (2011, p.34, 49, 52-53). Wolf (2011, p.34) interprets climbing the mountain as there being “(...) a single true morality, crystallized in a single supreme principle which these different traditions may be seen to be groping towards, each in their own separate and imperfect ways.” However, she disagrees with Parfit on whether this is true, and at times argues against Parfit with presenting an interpretation of the metaphor that fits with her claim that finding one supreme principle of morality is the ultimate task of moral theorizing, and that some tensions between Kantians and contractualists cannot be reconciled. As I interpret Parfit’s mountain, I will also include Wolf’s interpretation, and contrast the two when possible.

A plausible place to start is to think about our ordinary usage of the phrase ‘climbing a mountain.’ One can climb a mountain up to a point, or one can climb to the *top* of a mountain. Imagine two people, Alice and Bob. Bob is a wealthy amateur climber, and Alice is a successful professional mountaineer. Alice and Bob both go to Mount Everest. A month later, Bob returns and tells you he climbed Mount Everest. Knowing that summitting Mount Everest takes 2-3 months, and that Bob is not that good of a climber, you figure what Bob really means is he climbed Mount Everest to a point. 3 months later, Alice returns and tells you she climbed Mount Everest. Knowing that summitting Mount Everest takes 2-3 months, and that Alice is a professional mountaineer, you figure Alice means climbed to the top of Mount Everest. Nowhere in *On What Matters* does Parfit claim that three competing families of moral theories climb to the *top* of the mountain from different sides. This, we should take note of.

What would be the mountain moral theories are climbing, and what is atop of it? Intuitively, the mountain symbolizes the ethical discourse about what is the right thing to do, what responsibilities are, what grounds we have for believing or disbelieving our moral theories, and how should we act in light of these. Assuming the aim of inquiry, at least on a theoretical level, is truth, then atop of the mountain symbolizes moral truths that moral theories try to help us discover, and systematically account for at a general level. Surely, there are lots of moral truths at a particular level. It is wrong for me, right here and now, to kick a puppy for fun. We would like our moral theories to account for such particular moral truths too. If a moral theory implies that I can, right here and now, kick a puppy for fun, I would be reluctant to think that this moral theory is approximately true. The top of the mountain, however, is not an instantiation of other higher mountain tops. It is a singular point from which we can look at the landscape around us, and make systematic judgements at a general level. We might look towards East, see a forest, and claim that it is inhabited. Even if there are a few houses hidden within the forest, we wouldn't be wrong, *period*. Similarly, our moral mountain top would contain moral truths of a general level, which successfully captures and prescribes the wrongness of most instances of kicking puppies for fun. Often, we call such general claims "moral principles."

Moral principles need not always be *general*, whatever one takes generality to be. It is difficult to put forward a definition of generality, which could act as a threshold (Väyrynen, 2008; Hare, 1972). It is not straightforward to claim the following principle is not general enough for current purposes, for it would successfully capture and prescribe the wrongness of most instances to which the principle applies: 'It is wrong to treat French infants aged 2-4 as mere means if and only if affordable space travel is unlikely to be reached by 2050.' Nonetheless, we have an intuitive understanding of levels of generality of moral principles, and the type of general moral principles

our theories admit. Parfit surely has. Throughout his convergence project, Parfit does not *exclusively* concern himself with what a particular moral theory says is the right thing to do in a scenario. Rather, he is involved with finding intuitively general moral principles which seem to be true in the three rival theories. For example, The Consent Principle, according to which it is wrong to “treat anyone in any way to which this person could not rationally consent” (2011, p.181), and the Mere Means Principle, according to which it is “wrong to treat people merely as a means” (2011, p.212). These principles are not explicit statements of what we ought to do in a particular situation. Rather, they are intuitively general statements about the rightness or wrongness of classes of actions one might follow in a particular situation. Parfit’s search to find, say, optimific principles in contractualism seems to suggest that for Parfit, moral principles capture moral truths. Furthermore, they are kind of claims that we have access to, whose truth or falsity we can sometimes recognize. So, I suggest we put atop of the mountain moral principles of a general level, which successfully capture certain moral truths.

Wolf (2011, p.35) goes one step further, and claims that rather than general moral principles, atop Parfit’s mountain is “single supreme principle of morality” It is interesting to think whether this makes a difference for Parfit and Wolf, and the answer is not obvious to me. On the one hand, as we’ll dissect in more detail in throughout this work, Parfit thinks once you reach at the true general moral principles, it is merely a matter of conjoining them to reach the single supreme principle of morality (2011, p.413). This principle contains the properties that would describe a “single higher-level wrong-making property, under which all other such properties can be subsumed, or gathered.” (Parfit, 2011, p.414) On the other hand, it is plausible, or even likely, believe that if such a supreme principle exists, reaching it from previously discovered true moral principles is not merely a matter of conjunction. After all, if Parfit is correct and competing theories do in fact get

to same general moral principles, conjoining them to reach at the supreme principle of morality would have been done. This is not direct evidence against the thesis that supreme principle of morality, the top of the mountain, is not to be achieved by conjunction, but given the highly complex task of resolving recalcitrant tensions among seemingly different moral theories, it is not reasonable to expect that anyone who has the principles could make the conjunction. Arguably, something more than principles and the rational capacity to conjoin propositions is required. Perhaps this is why after a thorough criticism of Parfit's reconciliation of Kantianism and contractualism, Wolf chooses a more attainable mountain top. Parfit may have not shown the single supreme principle of morality but perhaps, he has gathered "a well-ordered set of principles" (2011, p.51). These principles would be necessary but not sufficient to discover, if it exists, the supreme principle of morality. Since we do not know whether such a principle exists, I believe it is reasonable to proceed with putting well-ordered general moral principles atop the mountain, which finds some support in Wolf's interpretation too.

One might think two claims I have made are in tension. On the one hand, I argued that the mountain's summit is the set of well-ordered moral principles which would completely describe our moral world. From the summit, we could successfully judge whether an act is right or wrong in virtue of the complete set of principles, and some reflection. I also stated that nowhere in *On What Matters* does Parfit claim that the competing moral theories reach the summit. Our best effort in moral theorizing so far, then, is more like Bob rather than Alice. If the supreme moral truth is evading us, then it is plausible to doubt my claim that we should put moral principles, as capturers of moral truth at a highly general level, at the top of the mountain. After all, according to the mountain metaphor, competing moral theories are not at the top; hence, they have not successfully captured the supreme moral truth ethical theorizing, from a moral realist perspective, aims to

capture. This need not worry us, because we can justifiably believe that some versions of Kantianism, contractualism, and consequentialism are not complete yet when it comes to capturing the supreme moral truth, if it exists. Arguably, for them to capture *the* moral truth, they would have to be complete. However, even if they are not complete, they would still get some of *the* moral truth. This point is also made by Wolf who argues that all three theories Parfit discusses are groping towards the top in “their own and *imperfect* ways.”(2011, p.34). Arguably, part of their imperfection stems from the fact that the competing theories do not contain *the* well-ordered set of principles that we put atop the mountain. There are possible cases wherein, since they lack the relevant moral principle which we would use to make and explain a true moral judgement, they would either not yield a judgement, or yield a false one. Nonetheless, they will, at the general level, be *approximately true* over the domain of all possible cases. As we hike through the mountain and as our ethical discourse progresses, we make modifications on the theories, which makes them more and more complete. As we do so they get closer and closer to *the* moral truth. I find this to be not only plausible but also supportive of the claim that moral principles occupy the summit of the mountain. That our current theories are not there yet leaves the possibility for progress. So, we can put the appearance of tension to rest, and carry on.

I have so far argued that the climbing of Parfit’s metaphorical mountain amounts to moral theorizing, at least the ones Parfit is concerned with, trying to capture moral principles, which themselves systematically capture moral truths. Now, let us turn our focus on the sides of the mountain. What would it mean for these moral theories to aim to capture the same set of moral principles, from different sides? A plausible starting point is to interpret the different sides as different paths. In this interpretation, Kantianism would take a different path to capturing, say,

The Consent Principle, compared to contractualism. But what does the path symbolize? And as Baumann (2018) rightly asks, why does which path we take atop not matter for Parfit?

Imagine a Kantian and a contractualist agree upon the moral truth of a principle. What remaining disagreement is there between them that would be captured by the different paths taken to reach that principle? If climbing the mountain symbolizes moral theorizing, perhaps it will be useful to look at moral theorizing to find an answer to that. Arguably, we do not want black-box moral theories which tell us what we ought to do. Rather, we seek theories which can also systematically explain why we ought to do what we ought to do. Put differently, we not only wish to know *what* matters, but also *why* it matters (Dietrich and List, 2017). Suikkanen (2014) argues that Kantians and consequentialists can agree upon intuitively right acts, yet disagree about what makes them so by giving competing explanations. Parfit acknowledges this disagreement, to some extent. He claims the disagreement among the experts are “deepest we are considering, not the wrongness of particular acts, but the nature of morality and moral reasoning, and what is implied by different views about these questions.” (2011, p.418-419) Leaping from explanations to moral reasoning is a big step, for moral reasoning can go well beyond explanations, and it would be a mistake to equate the two. However, coming back to the metaphor, I believe the explanatory disagreement, as well as Parfit’s comment give us an adequate ground for explicating what these paths mean. Parfit mentions what particular moral theories imply about moral reasoning, indicating that there is a particular way which, say, Kantian reason morally. Call this the Kantian framework. Within this framework, there is room for moral explanations as to why certain acts are right. They might, for instance, directly or indirectly appeal to some formulation of the Categorical Imperative to provide an explanation. That explanation can then allow us to get closer to grasping the truth of a moral principle, or in metaphorical terms, get us closer to the top. Of course, in the natural world,

we get closer to the top by walking the physical path. So, moral explanations given by competing theories is a natural candidate for the different paths taken by moral theorists.

Despite being a plausible candidate, taking moral explanations as what the path symbolizes is perhaps too narrow. In choosing a particular path atop a mountain, we do not only choose a physical path to walk on. We also choose what view we get as we go along. Take the Eastern path and you will have a view of the East. Take the Western path and you will not have a view of the East. There is a sense in which despite sharing their destination, *i.e.*, the top, our *experience* of climbing is partially determined by the path we choose. Imagine the mountain sitting in an interesting geographical point, which leads it to have sunshine on the Eastern path and gloomy clouds in the Western path.⁵ It is quite reasonable to claim that the experiences of two boulderers climbing from these different sides will also be significantly different.

Above, I have dubbed the way Kantians reason morally, which Parfit puts at the heart of the disagreement, the Kantian framework, and put moral explanations in this framework as part of different paths taken atop the mountain. However, as I have claimed, the different paths taken are not merely the different physical paths taken. They partially determine the climbing experience that goes beyond physically travelling from one point to the other in space. Similarly, the Kantian framework is likely to have a direct or indirect influence on the way in which a Kantian morally reasons, that goes well beyond the explanations given for the rightness of an act.

⁵ A quick internet search yields that this is in fact true in some parts of the World.

The question of what else exactly constitutes the Kantian framework is a tough one, and I admit I don't have a complete answer. There is a sense in which we can understand a Kantian way of thinking, which may figure in an agent's moral thinking. Employing this framework plausibly leads one to employ particular modes of deliberation, identify robust and reliable simple rules that one can employ without much contemplation, or single out relevant moral features of a particular case. Furthermore, it systematizes one's moral deliberation process. For instance, for Kantians, the Categorical Imperative acts as "a regulator of maxims of action" (Väyrynen, 2008). This responds to our demand that moral theories give us systematic verdicts and explanations, and not be black-box theories.

The Kantian framework is a heuristic guide, so to speak, that partially determines one's overall experience, assessment, and judgment of the moral world with its modes of deliberation. The same will hold for contractualism and consequentialism, and the frameworks that come with adopting them. Regarding the latter, Railton (1984; p.148) argues that "by itself, consequentialism is not a particular moral theory, but it becomes one, or a part of one, when combined with a theory of the good, a theory about which worlds are better or worse than which others." Similarly, Brown (2011) argues that consequentialism is like a *genus*, and particular consequentialist theories, like utilitarianism, are *species*. That which is a genus identifies particular properties which all its species would share. For instance, the consequentialist genus would defend maximizing a particular good and promote agent neutrality. I believe both Railton and Brown are making a move similar to one I'm making. What they refer to as 'consequentialism' is what I refer to as 'the consequentialist framework.' It is a heuristic framework for a particular way of moral deliberation, with particular explanations built into it. 'Consequentialists,' in Railton's and Brown's terminology, reason in a particular way proper to consequentialists. However, two consequentialists with differing

commitments as to what constitutes the good, despite using the same way, or method of reasoning, namely the consequentialist framework, can plausibly come to different conclusions.⁶ In metaphorical terms, it is not only the physical path, but also the view one sees when one takes that path, partially determining the experience of climbing. For example, in taking the consequentialist path, one adopts agent-neutrality as a general feature of one's moral thinking. They *see* and assess the moral world around them while conforming to agent-neutrality.

Adopting what I call the framework as the routes taken atop is similarly interpreted by Wolf (2011, p.52). Although she does not dive into the experience of climbing, and seeing the landscape that a particular landscape provides, she makes a very interesting point about how being committed to a framework can affect the path taken. So far, I have not made it clear whether the path atop is a single path from the bottom to the top, or whether it is a complex one with junctions involved. Parfit talks as if the paths are continuous from the bottom to the top, and exponents of moral theories discussed meet at the top, though he is not explicit about this. Unlike Parfit, Wolf talks of junctions on the road. Perhaps, the paths of Kantians and contractualist cross at certain points, yet they go on to take different ways. Being at the juncture, the Kantian framework perhaps leads a Kantian to take the road left, whereas consequentialist takes the road right for reasons I have given above. Wolf (2011, p.53) claims the path taken by Kantian shines so bright for the Kantian. The Kantian may even fail to realize there is another way available to them. The bright shining of a particular path is accounted by adopting the framework account I have put forward. If a path is

⁶ To avoid any confusions, I will keep using the framework terminology when I talk about the family of consequentialist theories, and will freely use consequentialism, and Kantianism and contractualism for that matter, to refer to particular theories Parfit endorses.

shining bright for someone who adopted a particular framework, then there must be something about that path which attracts the proponent of the framework. For example, at a moral juncture, a Kantian may simply see the path which prioritizes autonomy over optimificity, whereas the consequentialist may fail to even realize there is a path that goes through the prioritization of autonomy, given that they value optimificity above all else.

Now that I have presented my interpretation of Parfit's mountain, it is a good time to contrast this with Wolf's mountain. Wolf's mountain is how Wolf re-shapes the mountain metaphor that fits with her account as to what the aim of moral theorizing is, and the significance of disagreement among exponents of Kantians and contractualists. Wolf (2011, p.52) asks us to consider whether the aim of moral theorizing is to discover a single supreme principle of morality. If our answer to this question is yes, then Parfit's mountain and project is the best candidate of this job. However, Wolf argues that there is no good meta-ethical reason to assume the answer is 'yes' (2011, p.35). The moral world is a complex landscape, and there will be reasons for to consider different aims for morality. There will be times where aim of morality will clash with finding one supreme principle of morality. For example, can rationally prioritize effective moral decision-making procedures as our aim in moral theorizing, or articulating particular values fundamental to moral action. If these, and other plausible aims of morality, cannot be jointly realized to the maximum, then each meta-ethical position where we subscribe a different aim to morality will require a different mountain. As Wolf puts it, to realize all aims of morality to the maximum by getting the "most out of a moral theory, one must hike the whole range." (2011, p.53) Hence, if Wolf is right, there is not one but many mountains, atop of which is a particular aim of morality. Parfit's mountain can be in Wolf's mountain range, perhaps as the mountain with the aim of finding the

supreme principle of morality, but hiking Parfit's particular mountain, if I read Wolf correctly, is of low value when our aim is to alleviate the suffering of a particular group.

Although I sympathize to the thesis that we should be pluralist with respect to aims of morality, I do not see why this has to clash with Parfit's view and the mountain. Metaphorically speaking, the mountain top gives us an overview of the whole moral landscape, which includes cases where we would have different aims to employ moral theories. Granted, it is practically difficult to see the Eastern coast far away, and we would have a better understanding of the Eastern coast if we were on the Eastern coast. But same holds for the Southern coast, Western villages, and Northern mountains. In every intellectual endeavor, there will be practical difficulties in getting from a theoretical or abstract general point of view to a particular one. Nonetheless, I believe it is reasonable to expect to have this general point of view which would allow us, to the extent it can, to make sense of the coasts, villages and forests systematically. Being in the Western village, one might wonder why they decided to build a village there rather than somewhere else. However, from the mountain, we can see natural resources nearby, and make sense of it. Same holds for the moral landscape. Particular possible cases where we ascribe different aims to morality are merely part of the landscape, and we have a good grasp as to where things are in the landscape, and how they seemingly relate. It is with the assumption that there is a particular right thing to do and an explanation of why it is the right thing to do, which can *in principle* be derived from a well-ordered set of moral principles, we seek for the right thing to do in a particular cases. Contrary to Wolf, who sees no good reason to assume there must be just one mountain, I see no good reason to assume there must be many.

Let me re-iterate what I have said for so far. I began by claiming that Parfit's aim is to resolve the disagreement between experts which he deems as a threat to moral theorizing. It is with this aim in mind that we should be reading and interpreting Parfit. I then interpreted the elusive mountain metaphor. I argued that the mountain itself is the discourse of moral theorizing. Atop the mountain I put well-ordered moral principles. I acknowledge that this is perhaps controversial. One can be a particularist like Dancy (1983), and deny the truthfulness of moral principles, but still claim that there are moral truths. They might claim that there is a fact of the matter about the wrongness of torturing animals for fun, but subscribing to a principle that says 'It is wrong to torture animals for fun' will not help you get to that truth. My aim is not to defend generalism against particularism, and as such this falls beyond the scope of this work. Nonetheless, a particularist objection would miss the point of my interpretation of Parfit's mountain. I interpret the mountain with Parfit's aim in mind. That is, Parfit's mountain need not be *the* mountain. When it comes to resolving the disagreement in moral theorizing, Parfit's methodology is to revise and modify moral principles and what makes them acceptable. His continuous effort to revise moral principles, with the aim of getting to some moral truths, seems to indicate that Parfit thinks moral principles are adequate for getting to moral truths. I believe he takes them to be truth-apt, leaving the door open for their truth to be *grounded* by further moral truths. Naturally if one denies the truth-conduciveness of moral principles, like Dancy, then Parfit's project, in the way he constructs it, fails. Then, I have argued that the paths taken are heuristic frameworks that particular strains, families, or *genera* of moral theories come with. It is an open-minded way to view the moral world, to systematically and robustly observe, assess, and explain it. I have found some support for this claim in Wolf's account. Finally, I objected to Wolf's mountain (range) where she argues different aims of morality require different mountains. I claimed that different things we expect from moral theories are part of the

landscape and although it is impractical to investigate what exactly is required at particular points in the landscape, this is not a sufficiently good reason to overpopulate our moral landscape with mountains.

3.2: Convergence in Scientific Theorizing

Having analyzed and interpreted the mountain metaphor, let us turn our attention to the Convergence Claim. A good place to begin is with the word ‘convergence’ and how it is used by Parfit. Initially, Parfit uses the word ‘convergence’ to define the relationship between Kantian and Scanlonian versions of contractualism. His argument which identifies principles that no one could reasonably reject, *i.e.*, Scanlonian contractualist principles, with principles that everyone could rationally will to be universal laws, *i.e.*, Kantian principles, is dubbed The *Convergence* Argument (p.412). What is more interesting is that Parfit seems to think within this context, ‘to converge’ is synonymous with ‘can be combined.’ Right after dubbing his argument The Convergence Argument, he claims “Kantian and Scanlonian Contractualism can be combined.” (p.412) As we keep reading Parfit, we find another bit of textual evidence. Having *converged* or *combined* Kantian and Scanlonian formulations of contractualism, Parfit claims that the same argumentative structure which allowed him to do so was already applied to RC’ and Kantian contractualism. (p.412) As I exposed in section two, Parfit argues RC’ is implied by The Kantian Contractualist Formula, and proposes that these two theories “can be combined.” (p.411) This is another indicator that for Parfit, to converge two moral theories is to combine them. Finally, there is literature on the different convergence claims Parfit makes, and a liberal usage of convergence throughout. Baumann (2018) and Chappell (2021) use convergence to define Kantianism, contractualism, and consequentialism *morphing* or *combining* into TT. Perhaps more credence lies on the usage of the word by Scanlon (Parfit, 2011b, p.117), who uses the word ‘convergence’ at places

Parfit used ‘can be combined’ in his response paper that got published in *In What Matters Vol.2*. In sum, the textual evidence I have given indicates that both Parfit, and his critics and commentators use ‘to converge’ as ‘can be combined into.’

Parfit’s liberal usage of convergence and combination naturally gave commentators to use the same wording. This, however, I believe was a mistake on Parfit’s end. Given our interpretation of the mountain metaphor, I believe there are significant differences between two theories converging, and two theories being combined. This seems to have gone unnoticed by interpreters and commentators alike. It won’t be a stretch to claim that there is a lacuna, namely mistakenly interpreting ‘convergence’, and how it diverges from ‘to be combined’, in the Parfit literature, which we will aim to fill.

The aforementioned mistake perhaps is best emphasized by philosophers of science, who deal with scientific theories *qua* theories. Within the scientific domain, there is a clear delineation between combining and converging. To illustrate this point, let us take an example from history of physics. One theoretical virtue of scientific theories is that they retain previous results. Call this conservativeness. If conservativeness is upheld as a theoretical virtue, then one expects that a new theory B which replaces the old theory A successfully replicates the empirical results of A. Within the scientific community and the rational reconstruction of history of science, this virtue is surely upheld (Lakatos, 1971). Scientists in training are told that most new theories do not completely overthrow the old ones. Rather, they build upon them. They make the same predictions the old theories made, with higher accuracy, and then make some new ones. A fruitful illustration of this is found at the turn of the 20th century, when quantum mechanics was replacing classical mechanics. The new theory, being so odd and perplexing, suffered serious resistance from the

previous generation of established scientists. One argument in favor of quantum mechanics is that it *converged* to classical mechanics at the classical limit. Let us expand this claim.

Classical mechanics was developed to predict and causally explain classical systems, *i.e.*, system that were sufficiently large to the extent that forces at the atomic scale is negligible. If you're curious about the collision of two cars, there is no need to consider any energy loss that may happen at the atomic level, for it would be so small that it would have no significant contribution to understanding the dynamics of the collision process. If, on the other hand, you are curious about the collisions of two atoms, such atomic level effects cannot be neglected, and quantum mechanics was developed to better predict and explain such phenomena. But which theory should one use if one is looking at very small classical or very large atomical systems? What would happen at the intersection of the atomic and classical scales? It turns out that both theories yield approximately the same results at such scales (Shankar, 1994). It is then said that quantum mechanics *converges* to classical mechanics at the classical limit.

Does the convergence at the classical limit imply the two theories can be combined? Can we have a combined theory which divides its domain into two, and apply the appropriate sub-theory for the domain at hand? The answer is a strict no. The two theories are logically inconsistent, with different initial assumptions about the way the natural world behaves. Assuming that the two theories can be expressed semantically (Psillos, 1999; Chakravartty, 2001), their combination would be a contradiction of different sorts. They could, for instance, provide contradictory descriptions of natural world. While the quantum mechanics describes atoms as particle-waves, classical mechanics describes it them as point-particles. Due to this metaphysical commitment to the nature of the atom, these theories would also yield contradictory explanations regarding, say, atoms behave in collisions. The classical theory will say two atoms *indeed* collide and scatter,

according to quantum mechanics, two atoms can scatter without actually colliding. Moral of the story is that one cannot operationalize quantum mechanics with a classical worldview, or the classical *framework*, and vice versa. Similar to the consequentialist framework which is a heuristic guide to see and assess the moral world around us, the classical framework would be a particular way to observe and describe the natural world around us. Analogous to the differences between moral frameworks, there would be significant differences between adopting a classical or a quantum framework.⁷ Furthermore, their convergence would only indicate predictive similarities between the two theories. If we were to ask why we observe what we observe, they would yield inconsistent explanatory statements. Yet, this does not prevent the two theories converging at a specific range, about a particular aspect of the natural world, with emphasis on *particular*.

To sum up, rival scientific theories cannot be combined, for the end result would contain contradictions. However, they can converge into one another. Convergence of rival scientific theories means that in particular corners of their respective domains, they yield the more or less the same predictions as to how a system would behave. They would still contain rival and logically inconsistent explanations as to why the system behaves the way it does. However, since we are not combining the two theories, and merely taking the same predictions as indicator of the new theory's truth, we justifiably believe that they converge.

⁷ During the development of quantum mechanics, physicists adhering the classical framework tried to explain particular phenomena with classical mechanics, arguing that the relevant features for understanding the phenomena was classical at heart. As one can imagine, adopting the quantum framework had its own epistemic commitments when it comes to understanding and describing the phenomena. A good example of this period is the early atomic models. Here, understanding of the atom goes through a classical phase due to Rutherford, a semi-classical phase due to Bohr, and a quantum phase due to Schrödinger, Dirac, and Pauli.

3.3: The Structure of Moral Theories

Competing scientific theories can converge but cannot be combined. How does this help us in understanding what Parfit means by convergence? Plausibly, the moral domain is quite different compared to the scientific domain, and convergence in scientific theorizing need not be a guide to convergence in moral theorizing. Although this is *prima facie* likely, there nonetheless is something we can learn from the scientific domain. Notably, we can, following the structure of scientific theories, perhaps offer a broadly applicable structure for moral theories too.

In our example about the convergence of quantum and classical mechanics, note which parts of the theories at hand converge, and note which parts are logically inconsistent. They converged in their predictions, i.e., the content of scientific theories that we can somehow observe. Call this the *observable content* of scientific theories. Where the theories at hand diverge is the explanations they provide for the observable phenomena. While classical mechanics comes with its own metaphysics about particles and forces, quantum mechanics comes with a metaphysics its own. Importantly, they have widely diverging metaphysics of atomic particles. Classical mechanics treats them as point particles, whereas quantum mechanics treats atoms as particle-waves. Metaphysics of quantum mechanics is a rabbit hole of its own into which we will not dive, as it goes beyond the aims of this project. We only need to note that there are two metaphysical frameworks, that give rise to, among other things, two competing explanations. Let us call these metaphysical commitments the *unobservable content* of scientific theories. Two rival theories might yield the same predictions, that is they might be observationally equivalent, and yet contain different unobservable contents. They might posit different mechanisms explaining observable phenomena, or they might submit to different metaphysics. For instance, whereas Newtonian classical

mechanics treats space as absolute, Galilean classical mechanics, which is observationally equivalent to Newtonian classical mechanics, treats space as relative (Sider, 2020, p.57).

Dietrich and List (2017) argue that structurally, moral theories are analogous to scientific theories. Using a decision-theoretic framework, they represent moral theories using two parameters: “what matters and how it matters” (Dietrich and List, 2017, p.430). Their primary aim is to present a taxonomy of moral theories (Dietrich and List, 2017, p.422). What is important, however, is that this framework allows us to distinguish two parts of a moral theory. Let the *deontic content* of a moral theory be the verdicts about which acts are permissible and which are not. This content is the theory’s recommendations about guiding our actions (Dietrich and List 2017, p.424). Arguably, we do not just want black-box moral theories to tell us what to do. Our interest in ethics goes beyond how we ought to act. We often look for systematic explanations of why we ought to do what we ought to do. Ideally, we would like to identify some general principles that would underpin, and explain the deontic content. Let us call this part of a moral theory which explains its deontic content the *explanatory content* of a moral theory.⁸

Moral theories could, in principle, also output evaluative or axiological judgements in addition to deontic judgements. These would be judgements of the sort ‘It is better to ϕ rather than to φ .’⁹ Consequentialist theories can generate such outputs. Take, for instance, the evaluative judgement ‘It is better to kill one person rather than killing ten people’ In this work, we will not focus on such judgements and focus on deontic ones for two reasons. First, it is not clear whether all moral

⁸ Dietrich and List call this part *reason content*, and populate it only with rightness functions consequentializable theories admit. For reasons that will be clear later, I use explanatory content, and not appeal to rightness functions.

⁹ Compare this judgement with the deontic one ‘It is permissible to ϕ

theories, such as Kantianism, can generate evaluative judgements. Seemingly, a Kantian would say ‘It is impermissible to kill one person’ and ‘It is impermissible to kill ten people’ Yet, is not clear if a Kantian can make sense of one being judgement *better* than other. One piece of evidence supporting this claim is the literature on consequentializing. To ‘consequentialize’ a moral theory is to take a seemingly non-consequentialist moral theory, like Kantianism, and argue that it is in fact a consequentialist one. If Kantianism is consequentializable, then it can, in principle, generate axiological judgements. However, there is good reason to think not all theories are consequentializable(Brown, 2011, p.47). That a theory is not consequentializable is not sufficient to eliminate the possibility that it can generate axiological judgements. Nonetheless, for the purpose of this work, we can disregard such outputs of moral theories in favor of deontic verdicts because as Baumann (2018, p.209) notes, Parfit’s argument is aimed at resolving “deontic quarrels” Indeed, Parfit often talks of an act being ‘wrong’, ‘allowed’, ‘disallowed’, rather than acts being ‘better’ or ‘worse’. So, we set non-deontic judgements aside, and proceed with deontic verdicts as the main output of moral theories.

What would the explanatory content of a moral theory consist of? Above, I have argued that we demand our moral theories systematically capture particular moral truths about what we ought to do. These moral principles can also figure in the explanations of what we ought to do in a particular case. Imagine our professional mountaineer Alice is being nice to wealthy Bob. When Bob asks Alice if she wants to catch a movie or grab a bite, she always says yes. Unbeknownst to Bob, Alice actually does not like spending time with him. Her motivation in doing so, however, is the simple expectation that Bob will cover Alice’s expends in her upcoming summiting trip to Mount Kilimanjaro. Our intuition is most likely that what Alice is doing is wrong. She is deceiving Bob by pretending to be his friend when, in fact, she is not committed to whatever makes one a genuine

friend. She is doing this because she hopes for material gains, in the form of her expenses being covered. Put more generally, she is using him as a mere means. Our expectation is that an approximately true moral theory will be aligned with our intuition that what she's doing is wrong. If that's the case, such a moral theory will generate the verdict 'It is wrong for Alice to pretend to be a friend of Bob'. If we ask why that is the case, more often than not, the answer, *i.e.*, an explanation as to why the verdict is true, will involve a moral principle. For example, the explanation can be 'Because it is wrong to deceive people', or 'Because it is wrong to treat people as mere means'. This should not come as a surprise, since *general* moral principles are Humean-like general statements or Humean principles, *i.e.*, they are metaphysically idle yet practically useful generalizations of wrong-making properties (Enoch, 2019, p.2), and when faced with a particular instantiation of the general statement, appealing to the general statement as the *explanans* is a natural move.¹⁰

One could argue that using moral principles in the above manner as explanans is not explanatory at all; that used as such, moral principles are explanatorily idle. Such a criticism has been raised by Berker (2019). Berker argues that the rightness or wrongness of an act is determined not by the moral principles it instantiates, but rather the moral properties it has which the moral principles are generalizations of. In the case of Alice and Bob, it is not the case that Alice's act is wrong in virtue of it violating The Mere Means Principle; rather, it is in virtue of the fact that the act itself contains the moral *property* of treating people as mere means. Berker claims since Humean-like statements, *i.e.*, moral principles in our case, are merely generalizations about particular moral properties, and since wrongness of the act is in the property itself, the principle provides no

¹⁰ Compare this to the following example from the natural domain: if one were to ask why a particular person is not hatching an egg, one would say 'Because humans are mammals, and mammals do not hatch eggs'.

explanation. There are two ways to meet Berker's challenge. We will now go through the first way, and cover the second way later when we exhaust the explanatory content of moral theories.

Berker's challenge revolves around the notion of *grounding*, an explanatory, non-eliminative, and non-reductive relation (Fine, 2001, p.15). We can respond to his challenge by arguing that the explanation we demand from our moral theories is not always similar to grounding. Of course, Berker is not simply conflating two distinct notions. Depending on one's preferred epistemology, an explanatory chain, viz., a series of statements in the form '*p* because *q*', '*q* because *r*', '*r* because *s*'..., can end at some point, '*y* because *z*', or it can circle back to the initial *explanandum* '*y* because *p*'. We can call these two foundationalist and coherentist epistemologies respectively. Berker is operating with a foundationalist epistemology. So, his aim is to get *the* fact of the matter which fully explains the *explanandum*. However, even if Berker is correct, the fact that moral principles cannot figure as *the* fact of the matter does not entail that they do no explanatory work.

What constitutes a good moral explanation is a tough question that goes beyond the scope of this work.¹¹ Nonetheless, like any other explanation, "the adequacy of a (moral) explanation depends on context" (Enoch, 2019, p.26). In certain contexts, appealing to The Mere Means Principle can be explanatory for a certain moral agent. Imagine Charles is presented with the verdict 'Alice's action toward Bob is morally wrong', and Charles asks 'Why?' Our initial answer to Charles, following Berker, can be 'Because Alice's action has the property of treating people as mere means'. Then, Charles is faced with a similar verdict, wherein Dennis treats Edmund as Alice treats Bob. The verdict goes 'Dennis' action towards Edmund is morally wrong', and Charles once again asks

¹¹ See Majors (2007) for an overview.

‘Why?’ What seems to be at work in such a scenario is that when Charles is presented with an explanatory statement in the case of Alice and Bob, it is plausible that he assesses the wrong-making property to be that *Alice* treats *Bob* as mere means, but fails to notice that all such instances of X treating Y as mere means is wrong in virtue of treating people as mere means property. So, he is once again in need of an explanation in Dennis’ case.

Charles can be forgiven for thinking that it was only wrong for Alice to treat Bob in the way she did, in the particular case constructed, and not see why it was the case that it was wrong for Dennis to treat Edmund in the way he did. What if we gave Charles an explanation appealing to The Mere Means Principle? The principle will systematically deem all acts which contain the property of treating people as mere means *prima facie* wrong; this systematicity was perhaps lacking from the initial explanation which appealed to the property which the principle generalizes. So, when Charles asks ‘Why is Alice’s action towards Bob is wrong?’, we answer ‘Because it violates The Mere Means Principle, according to which it is *prima facie* always wrong to treat people as mere means’. It is not that Alice has any moral importance insofar as the wrongness of the act is concerned, since it would *prima facie* apply to all moral agents acting in the way Alice did. The moral principle-involving explanation then aids Charles to better understand what was wrong with Alice’s action in a way that he can systematically apply to all future cases. Then, it would be plausible for Charles, when faced with Dennis’ treatment of Edmund, to say ‘Ah! Dennis’ treatment of Edmund is wrong *because* it violates The Mere Means Principle.’ As long as Berker, or an ally of his, cannot rule out such a pragmatic role for moral principles, we cannot deny moral principles do some explanatory work.

The explanatory content of moral theories will then contain moral principles which can do explanatory work. Is there anything else that could be a part of this content? Can we ask for an explanation, for instance, as to why a moral principle is true, or what grounds we have for believing a moral principle, or why we take it to capture some moral truth? Recalling the explanatory chain presented above, is there a proposition z which can potentially end the chain, of which no further explanans can be given? To be explicit, two different types of explanation are at work here. The first type, call it explanation (1), concerns explanations as to why a particular verdict generated by the theory is true. For that, I argued moral principles can play an adequately explanatory role. This type of explanation will plausibly be practically employed in our everyday theory-informed moral decision-making. A more abstract type of explanation, occurring at a theoretical level can be about the principles employed in explanation (1). Let us call explanation (2) to be an explanation as to why a moral principle is true. The final step in constructing the sketch of the structure of moral theories is then to find what could play such a role.

To answer that, let us imagine we are a consequentialist who endorses The Mere Means Principle, according to which it is wrong to treat people as mere means. We are faced with the challenging task of finding a relation between The Mere Means Principle, and some proposition z , which would fully explain the truth of the principle, and itself can admit of no further explanation. From the consequentialist framework, what would justify endorsing this principle? The answer, I claim, is some sort of maximizing the good claim that underpins, and sometimes even defines, consequentialist theories. Put differently, something along the lines of ‘The Mere Means Principle is correct because it is optimific’ or ‘That all instances of The Mere Means Principle admit of optimificity property’ is at play.

To see why an optimificity-involving proposition would explain The Mere Means principle, let us assume a Kantian also endorses the same principle. So, the two are in agreement regarding the *prima facie* verdicts that are generalized by the principle, and some explanatory work done by the statement. What dispute is left between the two? As Baumann (2018) notes, why, in the case of a particular verdict explained by The Mere Means Principle endorsed by both the Kantian and the consequentialist, give competing explanations?

This brings us closer to Berker's challenge (2019), and the second way we can meet it. Of concern is what *grounds* The Mere Means Principle, and it is the grounds that ground the grounded wherein the disagreement among Kantians and consequentialists remains. Kantians are likely to claim that the principle is true *because*, viz., it is true in virtue of, some proposition *z* which involves a Kantian concept like being rationally willable. In contrast, the consequentialist's claim is likely to involve a concept like optimificity. It is important to note that the difference between the two, and the contractualists too for that matter, is not "merely notational variants" (Baumann, 2018, p.200). Plausibly, the Kantian will have an intuition such that employing being rationally willable "is just too different" (Enoch, 2019, p.29) than employing optimificity to ground the The Mere Means Principle. That is, it will seem like the tension between the two is not resolvable even in principle, and a Kantian ground could not cease to present a competing grounding claim "without losing a constitutive feature" (Baumann, 2018, p.200). Let us call these propositions which competing moral theories present as grounds for the moral principles they endorse 'the grounding claim'.

To illustrate this point further, we can recall Wolf's (2011, p.53) interpretation of the paths taken atop the mountain, where she argued that at a junction, one path might be shining so brightly for the Kantian. Imagine a Kantian and a consequentialist are at the metaphorical junction of

providing a grounds for The Mere Means Principle. For the Kantian, the path which employs being rationally willable shines so brightly. In that path, trees are greener, birds chirp more beautifully. When they take a look at the other path, they are likely to think that the consequentialist path is so different from their own, they could not, without letting go of some constitutive facts which makes the Kantian path so bright, choose to take the consequentialist path. The very same facts, which make the Kantian path so bright and attractive for the Kantian, make the consequentialist path dark. Because for the Kantian, it is being, for instance, being rationally willable which makes the path so bright, and it's the consequentialist path's lack of being universally willable that it, as a path one can take at a juncture, makes it so different for the Kantian.

The disagreement among moral theories when it comes to their grounding claim is not merely semantic. As moral agents, our experience of assessing the world around us also is not in agreement. To see why, one more time visit the mountain metaphor. I have argued that the route taken atop from different sides as the frameworks Kantians, consequentialists, and contractualists phenomenally employ to see and assess the moral world around them. These frameworks are the irreconcilable ways of experiencing morality, seeing optimificity versus being universally willable as *the* right-making property of actions, even if we can come to an agreement with one another as to whether an act is indeed right or wrong. The grounding claim competing moral theories admit is an expression of this phenomenal difference. It is a way to reduce and convey the very different experiential *thing* about being a Kantian or a consequentialist that is hard to convey; but one must try, nonetheless.

Before ending this section, let us revisit Berker's challenge, who claims principles can do no grounding work. According to Berker, 'The Mere Means Principle is grounded in the principle of

optimificity' would be false. Parfit, however, speaks of principles having properties (2011, p.413), and properties can ground. We could then interpret principle-involving grounding claims as being convenient expressions of generality of wrong-making properties. The claim 'The Mere Means Principle is grounded in the principle of optimificity' is then simply a particular and convenient way of saying 'The Mere Means Principle is grounded in the fact that every possible action which does not violate it has the property of optimificity.' Although the metaphysics of grounding is best not ignored, Parfit seems less concerned with it than resolving the disagreement among experts. It might be argued that leaving the grounding discussion for later, once there is considerable agreement, is a strategy one might take. Nonetheless, unlike the first way of meeting Berker's challenge, where we explicitly referred to principles to do some grounding work, here, we can simply concede and use property-involving language to ground the moral principles.

Let me end this section with a summary of what I have said. In the previous section, I argued that convergence in science involves making distinctions between different parts of scientific theories; that the structure of scientific theories informs the convergence dispute. Following the framework proposed by Dietrich and List (2017), I separated the structure of moral theories at hand into two: deontic content and explanatory content. This, I believe, should not be controversial for anyone who thinks moral theories should tell us what to do, and why to do it. The two contents give answers to these two questions respectively. The deontic content is the theory's verdicts. Things got more interesting with the explanatory content. Analogous to the unobservable content of scientific theories which underpins the observable content, our demand that moral theories go beyond being black-box generators of verdicts led us to put moral principles as doing some explanatory work as to why a particular verdict seems right. However, the disagreement among experts of moral theorizing is perhaps deepest as to why particular verdicts are right or wrong (Parfit, 2011, p.418-

9). It seems unlikely that the disagreement can be reduced to competing moral theories employing mutually exclusive moral principles, nor is this aligned with Parfit's work which revolves around finding shared principles between Kantianism, consequentialism, and contractualism. The remaining disagreement is about what ultimately determines the truth of both verdicts, and principles; namely, the grounding claim. The competing theories irreconcilably contain different grounding claims. This, I claimed, is aligned with our interpretation of Parfit's mountain. The different routes taken atop do matter, for they give us different experiences of climbing. Analogously, being a Kantian or a consequentialist give us different ways of experiencing the moral world, and all the claims they make about the moral world would be grounded by the grounding claim these theories admit.

Chapter 4: The Narrow Convergence Claim and TT

Having gone through the structure of moral theories, let us get back to Parfit and CC. In section 3.2, we took a look at what convergence means in the philosophy of science context. With the explicit structure of moral theories presented, we now can apply distinctions in philosophy of science to Parfit, and find some textual evidence to support this move.

First of all, recall that convergence is not combination. So, if CC is to obtain, it need not be the case that Kantianism, contractualism, and consequentialism can be combined into one another. Recall that in science, it is the observable content of scientific theories which converge to one another. Analogously, we can expect the deontic content of moral theories to converge. Let us then initially define The Narrow Convergence Claim as follows:

The Narrow Convergence Claim (NCC): Kantianism, contractualism, and consequentialism yield the same verdicts for all real and possible cases.

If NCC is true, then the argument from disagreement loses some of its force right away. The claim implies that there is in fact considerable agreement among ethics experts, that we are in a position to know what we ought to do irrespective of one's moral-theoretical commitments. Even this, to some extent, is sufficient to grant Parfit that he has partially succeeded in his aim to reduce disagreement among experts. However, throughout *On What Matters*, Parfit does not primarily concern himself much with particular verdicts.¹² Plausibly, this is because Parfit expects moral

¹² To be clear, Parfit does consider cases. However, his primary aim in doing so is not to find out the right thing to do. Rather, he tries to show one can arrive at the right thing to do starting from moral principles, and then shows how one can get to said moral principles from competing moral theories.

theories to systematically yield verdicts; a sentiment I am sharing. For Parfit, a verdict is true if and only if it is allowed by some true principle (2011, p.412).

According to the Humean picture of principles under which we have thus far operated, moral principles are generalizations of particular wrong-making properties. So, if a verdict is justified by, say, The Mere Means Principle, it then implies that the verdict is true because it contains the wrong-making property of using people as mere means. Furthermore, the property of using people as mere means is only captured by The Mere Means Principle.¹³ Any theory Parfit concerns himself with which successfully captures the wrongness of an act in which the using people as mere means property is present would also contain The Mere Means Principle in its explanatory content. Then, three competing moral theories which obtain convergence in their deontic content also obtains partial convergence in their explanatory content. NCC, I have above stated, was perhaps not up to the standard Parfit was striking for, nor was it following the primacy of moral principles in Parfit's thinking. Yet, we have now found how moral principles can be figured in NCC, for the verdicts are to be *underpinned* by them. Let us then reformulate the claim as:

The 'Narrow Convergence Claim' (NCC'): Kantianism, contractualism, and consequentialism yield the same verdicts, and they get to those same verdicts by using the same moral principles.

¹³ This assumes we are striking for simplicity when it comes to creating our principle ontology. Of course, one could create a new principle by disjoining two different principles, in which case the new principle would also capture the property of using people as mere means. I see no constructive reason to commit to such an ontology.

Alternatively, we can formulate it as:

The Narrow Convergence Claim” (NCC”): Kantianism, contractualism, and consequentialism contain the same moral principles, and they yield the same verdicts in virtue of aforementioned moral principles.

If NCC” is true, then Kantians, consequentialists, and contractualists agree upon what we ought to do in a particular situation, and they agree that what we ought to do in a particular situation follows by one or more moral principles they all share. So, if NCC” is true, then, *ideally*, one would expect there to be *considerable* agreement among experts in moral theorizing. If there is considerable agreement among experts, then the argument from disagreement is dissolved.

One could argue that there are too many if's involved, that this is not the *actual* case. In fact, some *actual* moral experts disagree with Parfit from the get-go with regards to his versions of Kantianism, contractualism, and consequentialism (see Herman, 2011b; Scanlon, 2011b). They might think that in the particular way Parfit constructed Kantianism, contractualism and consequentialism, they do yield the same verdicts but Parfit's construction of these theories is wrong to begin with, that his Kantianism is not a plausible reconstruction of Kant's thought. These will be people who Parfit fails to convince, and nothing more can be said about them or the disagreement they have with Parfit. Perhaps a more interesting class of experts are people who are convinced by Parfit's construction of the three theories, yet continue to disagree; for example, those who fail to apply the theories in ideal ways. It is plausible to imagine Parfit's work has convinced some people that these three theories are aspiring for the same moral truths in their own ways, but these people fail to apply those theories correctly. They would, for instance, despite using the same principles, wouldn't be able to get the same verdicts. Arguably, this class of disagreement among another group of *actual* experts should not worry Parfit. It is an issue other than Parfit's account that leads them to

disagreement. It is plausible to expect these people should, *ideally*, be in agreement, and we can expect them to, with reflection, actually agree. I believe Parfit's aim in presenting TT is to acknowledge the *actual* disagreement. He puts it forward as a way to *possibly* dissolve it. If you believe that TT is true, then, ideally, with enough reflection, you will get to realize that there is *actual* agreement among you and other experts. Same would hold for NCC", since it is a weaker version of TT.

Let us get back to TT, and the way Parfit prefers to formulate convergence. In his words, TT is expressed as "TT: An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable." (Parfit, 2011, p. 413). On the surface of it, there does not seem to be considerable difference between NCC" and TT. While NCC" states there is a finite set of moral principles, TT states a particular property of those principles, namely that they are optimific, uniquely universally willable, and not reasonably rejectable. In principle, this could guide us get to those principles in a way NCC" does not. Unfortunately, it is not the definition of TT that is problematic. Rather, it is Parfit's liberal usage of 'combining' competing theories into one another during his argumentation to get to TT. There are more problematic issues to unpack behind TT, such as combination leading to competing explanatory statements being present in TT, which NCC" does not have.¹⁴ Furthermore, the guide to get to the principles needed may clash with the grounding claims of Kantianism, consequentialism, and contractualism which led commentators like Baumann (2018) to think the final product, i.e., TT, is unsustainable.

¹⁴ As I have argued above, competing theories are likely to give competing explanations regarding a particular case or a principle.

It will be useful to employ the picture about the structure of moral theories presented in the previous section to better expose the difference between NCC'' and TT. With regards to convergence of their deontic content, the two claims are on par. When it comes to explanatory content of moral theories, I have argued that there are two distinct elements populating this content: moral principles, and grounding claims. NCC'' simply tells us that there is a finite number of moral principles which could allow us to capture the truth of the verdicts of a theory. Furthermore, they will do some explanatory work when it comes to why the verdicts seem true. However, NCC'' is likely to remain silent if we were to ask what *grounds* the fact that a given principle it endorses is true. Depending on what one's aim is, this may or may not present itself as an issue. If one is merely aiming for considerable agreement among experts, like Parfit does, then one need not worry about the lack of clarity regarding the grounds. In terms of the metaphorical mountain, it can be a point of intersection of the different routes taken above, where the three competing theories meet. Recall that the different routes taken is to be interpreted as the way one experiences the moral world around us. Having taken different routes, despite being in the same spot, the way a Kantian and a consequentialist assess the landscape around them will nonetheless be different. They might, for example, agree that there is a forest to the West. They might also agree that there is a forest to the West because there is a sufficient number of densely packed trees to the West; but disagree as to whether it is a pine forest, or a birch forest.

Let us contrast NCC'' with TT with regards to convergence in the explanatory content. Similar to NCC'', TT has convergence into a finite set of principles. However, notably different from NCC'' is that it involves concepts of optimificity, being universally willable, and being not reasonably rejectable. These, as I stated in section 3.3 are the grounding claims of competing moral theories. When asked why a particular moral principle seems to be true, the consequentialist will say '*Because,*

it is optimific' or 'Because it is grounded in it, or all of its instantiations, having the property of optimificity.' If grounding claims are also claimed to converge, what we have is complete combination. TT then combines Kantianism, contractualism, and consequentialism in their totality. However, as stated above, there is a strong sense in which the grounding claims of these competing theories are incompatible, that they cannot be reasonably sustained together without losing their constitutive features. Unlike NCC'', which can be upheld by Kantians and consequentialists, TT cannot, for Kantians are likely to deny that what grounds the principles of TT have anything to do with optimificity, and similar objections will be raised by consequentialists and contractualists. Although Parfit himself is not concerned with what would ground TT, this is not sufficient reason to park these concerns aside. If TT, as Parfit claims, is the closest thing we have to a supreme moral truth, we should be able to not only ground it, but get insight from it about the aims and nature of morality. As such, TT is a much more demanding plea for agreement, one that is likely to fail, given the deep disagreement among experts when it comes to the nature of morality (Baumann, 2018; Parfit, 2011). It is worth noting that we have seen this phenomenon in scientific theories too. Although convergence at particular domains is possible, total combination leads to a combined theory which implies a contradiction. Quantum mechanics and classical mechanics cannot be reconciled without either one of the theories losing their constitutive features about the nature of atomic particles, for example. Although scientists working with quantum and classical frameworks can agree upon the two theories converging at particular domains, they historically have not agreed upon the metaphysical commitments of the opposing party.

The metaphor that keeps giving also provides us with reasons to doubt TT can be genuinely adopted as a moral theory, from the perspective of a moral agent. Unlike NCC'' which is a point of intersection of different routes, TT seems to require us to have climbed all three routes at once.

This is because the grounding claims, of which NCC'' admits none, are the semantic expressions of our phenomenal experience of climbing the mountain from one side. To experience and assess the world as a Kantian is to appeal to some form of universal willableness, some formulation of The Categorical Imperative, the concept of duty, etc. The difference between being a Kantian, and seeing the moral world around us from the Kantian framework is, I claimed, is different than being a consequentialist, climbing the consequentialist route, exposing oneself to different aspects of the moral world. It is, after all, not just semantics.¹⁵ Kantians are not consequentialists, and the deepest of disagreement among them is to be found in their diverging opinions as to *why* the things that matter do matter. TT admits of the three separate grounding claims of competing theories, which we can interpret as needing to adopt Kantian, consequentialist, and the contractualist frameworks. Metaphorically, it requires us to climb all three routes at once. This, as one can imagine, is not possible. To answer Baumann (2018, p.198), who asks why do the different routes taken atop do not matter, we can firmly say it matters.

At this point, we can ask why Parfit is committed to TT, a much more demanding criterion for agreement among experts, rather than something along the lines of NCC''. To either dissolve or reduce the argumentative force of the argument from disagreement, which is Parfit's primary aim, NCC'' would suffice. Agreeing upon which acts are right and wrong, and which principles would constitute a minimal ontology for picking out the set of wrong-making properties which makes wrong acts wrong would in itself a huge accomplishment indicating the progress in moral theorizing, and perhaps if left at NCC'', the legacy of Parfit's success in reconciling Kantianism, contractualism and consequentialism would be viewed differently.

¹⁵ As Baumann says (2018, p.200), competing theories can sometimes even define themselves as the opposition of other.

I believe Parfit genuinely thought TT obtained. He seems to think he has erased the explanatory disagreement that NCC'' leaves room for. Plausibly, this is due to his principles-first approach to moral theorizing. Throughout *On What Matters*, Parfit employs principles to generate verdicts, e.g., it is wrong of Alice to treat Bob in the way she did because it violates The Mere Means Principle, and that the truth of the claim 'It is wrong of Alice to treat Bob the way she did' is grounded by The Mere Means Principle itself. That a complete set of principles, on which the competing theorists can agree upon, would yield an approximately true theory of morality with explanations built into it is what Parfit aims for. Furthermore, he believes he succeeds in this effort. At the end of volume 1, he says:

(...) disagreements are deepest when we are considering, not the wrongness of particular acts, but the nature of morality and moral reasoning, and what is implied by different views about these questions. (...) It has been widely believed that there are *such* deep disagreements between Kantians, Contractualists, and Consequentialists. That, I have argued, is not true. These people are climbing the same mountain on different sides. (2011, p.419)¹⁶

Having claimed that TT is true, it is plausible to think Parfit did not want to settle for a weaker claim like NCC''. That he thinks the 'deepest disagreements about the nature of morality' are resolved by TT seems to hint that Parfit has a particular conception of theory-driven moral explanations, and that principles, whose convergence he successfully argued for, were sufficient.

¹⁶ Emphasis mine.

This makes it all the more interesting that in volume 2 of *On What Matters*, he claims that deontic equivalence is sufficient for defending his view:

Some other moral disagreements are not about which acts are wrong, but about *why* these acts are wrong, or what makes them wrong. Different answers are given by different systematic theories, such as those developed by Kantians, Contractualists, and Consequentialists. Such disagreements do not directly challenge the view that we are able to recognize some moral truths. In defending this view, it is enough to defend the claim that, in ideal conditions, there would be sufficient agreement about which acts are wrong. (Parfit, 2011b, p.554)

Here, Parfit seems no longer committed to his claim that explanatory disagreements are resolved by TT, in virtue of having climbed atop the mountain where only agreement on principles reached. Due to the scarcity of textual evidence, it is difficult to say why Parfit committed to TT, or what exactly he thought of it. My aim in this project was to interpret convergence in terms of Parfit's aim to resolve the disagreement, the structure of moral theories, and our experiences of the moral world as Kantians, or consequentialists, or contractualists. In light of these, not only I believe TT is not true, but is also is not genuine moral theory that would fit with *being* a Kantian or a consequentialist or a contractualist.¹⁷ Even if Parfit managed to show agreement on principles and verdicts, the experience of being a Kantian and to have a framework, and assess the moral world around oneself evades TT. One can only take one route atop the mountain. TT is a point where

¹⁷ I'm only mentioning Kantianism, contractualism and consequentialism as these are the traditions Parfit and we have worked with. It is not an exhaustive list of moral theories. One could surely put, for instance, virtue ethics there too. Given our aim in this thesis to interpret and understand Parfitian convergence, we have not focused on other *genuine* moral theories.

moral theorists can reach, however it has no particular route belonging to itself. Upon arriving the summit via two different routes, a Kantian and a contractualist will still be a Kantian and a contractualist, and provide their different explanations as to why the deontic verdicts they agree upon are true. These will be provided based on their journey above, as moral agents who experience the moral world in different ways. By virtue of lacking this feature, TT is merely a sterile point in the moral space, where moral facts may lie, but questions as to what makes such facts true is left unanswered.

Nonetheless, one can, using Parfit's own work, make a weaker claim like NCC" which I argue to obtain in light of the conditions I stated. That Parfit failed to notice that TT does not obtain should not diverge our attention from the fact that he arguably did succeed in his aim. If NCC" is correct, there is considerable agreement among experts when it comes to what we ought to do. Furthermore, there is further agreement as to what explains the rightness or wrongness of those actions. Surely, as explanation is context- and agent-sensitive, there will be a point where competing theoreticians will give incompatible explanations, *viz.*, provide grounds for what makes right acts right. Is that a problem? I don't know. It is possible that a pluralism with respect to wrong-making is required. It is possible that disagreement is in fact a good thing, for it forces us to be open-minded, and revise our beliefs (Chappell, 2021, p.78). One net positive I stand by is that if NCC" is correct, then we would be in a position to know which way our current debates should move.

Chapter 5: Underdetermination of Moral Theories

This study of Parfitian convergence led us to interesting distinctions about the structure of moral theories, moral explanation, and grounding of moral truths. Throughout the way, we occasionally made analogies with philosophy of science, and scientific theorizing. In the same vein, Baumann (2018) raised an interesting criticism against Parfit that differs from much of the existing criticisms. The majority of the existing critics focus on Parfit's construal of his preferred versions of Kantianism, consequentialism, and contractualism, and thereby rejecting convergence, or TT, based on, for example, contractualism not being properly portrayed (Scanlon, 2011b). Baumann, on the other hand, assumes that TT obtains, and raises an intriguing argument often deployed by philosophers of science: underdetermination of theory by evidence. In this last section, I want to investigate Baumann's argument, raise a criticism against it, and explore whether it applies to NCC". The argument goes as follows:

P1. If two moral theories (MT) can account for exactly the same evidence, it is equally reasonable to believe either of them.

P2. If it is equally reasonable to believe either of two MT, we have no reason to attribute truth to one but not the other.

P3. If two MT contain incompatible propositions, they cannot both be true.

P4. If two MT cannot both be true, and we have no reason to attribute truth to one but not the other, then none of them should be considered true.

UMT. There are alternatives to even our best moral theories that can account for exactly the same evidence while containing incompatible propositions.

C. Therefore, even our best moral theories should not be considered true. (Baumann, 2018, p.208)

The argument, applied to Parfit, then claims that, assuming TT obtains, Kantianism, contractualism, and consequentialism, and perhaps even TT itself, all account for the same evidence. So, we have no reason to believe which one among them is the one true moral theory.

As it is the case in the underdetermination literature in philosophy of science, a lot depends on what we take moral evidence to be. Following Dietrich and List's (2017) pioneering work, Baumann (2018) proposes to ascribe verdicts the role of evidence in ethics.¹⁸ However, there is some unclarity with regards to which verdicts are to play this role. At different points, he talks of taking *particular* verdicts as evidence in ethics:

- that the *particular* verdicts are epistemically on a par with the data of scientific theories (p.207)
- What the analogy presupposes is only that *particular* verdicts have to be accounted for by moral theories (p.207)
- ascribing *particular* verdicts the role of evidence in ethics (p.207)

Other times, the 'particular' qualification is removed, and he speaks of verdicts of a theory:

¹⁸ Baumann swiftly acknowledges that verdicts need not exhaust moral evidence, and right after presenting the argument, he goes through additional types of evidence available for arguing against underdetermination.

- treating both the data of scientific theories as well as the deontic verdicts of our moral theories as extensions of those theories. (p.207)
- it is the theories' deontic verdicts alone that take the place of empirical evidence. (p.209)

The 'particular' qualification carries an importance, in that its removal leads to a problem of circularity. On the other hand, keeping the qualification raises the question 'Which verdicts?' Based on what criteria should we pick the particular verdicts that will play a role? Assuming we settle these questions, is NCC'' also underdetermined?

5.1: Moral Evidence

Let us start with taking unqualified verdicts as evidence in ethics. Assume there is a moral theory, Anti-Utilitarianism, according to which we ought to maximize suffering. A verdict that will follow from this theory when you see a puppy is 'You ought to kick this puppy', for it will maximize suffering.¹⁹ If all, viz. unqualified, verdicts of a theory are taken to be evidence, then, I should take the verdict that I ought to kick a puppy as evidence that Anti-Utilitarianism is true. Furthermore, as a rational agent, I should be responsive to that evidence. One way I can respond to this piece of evidence is to kick the puppy. Clearly, this is very absurd, and I do not think Baumann commits to such a view about moral evidence. However, it does highlight that if verdicts are to assume the role of evidence, then there should be a qualification over them. Not all verdicts should play this role. As Baumann says at points, it is *particular* verdicts to which we should ascribe the role evidence in ethics.

¹⁹ Assume this is the way I can maximize suffering in the World at this particular moment.

The question then is how to find a criterion of particularity. Based on what can we separate evidently false verdicts of Anti-Utilitarianism from evidently true verdicts of other moral theories? Unfortunately, neither Baumann (2018), nor Dietrich and List (2017) give us a criterion that could allow us to demarcate between evidently false and evidently true. However, we could, like the authors themselves, look at philosophy of science, and perhaps find useful distinctions about what confirms scientific theories, and what scientific theories themselves generate.

Dietrich and List (2017) talk about the observational equivalence of competing scientific theories, which leads them to think about the deontic equivalence of moral theories. How do scientific theories become observationally equivalent? Here is a rough sketch that is adequate for our present purposes:

1. Scientific theories generate predictions.
2. Predictions get tested.
3. Once a prediction is confirmed, it, i.e., the confirmed prediction, takes up the role of evidence in favor of a theory.
4. If competing scientific theories generate the same predictions, and they are equally confirmed, they are observationally equivalent.

So, the observational content of scientific theories is not always evidence in favor of the theory. History of science is filled with failed predictions. Rather, the observational content is the predictions, and the evidence is confirmed predictions. Put differently, *particular* predictions are evidence in science. Let us sustain the analogy between moral and scientific theorizing, and present a rough sketch for moral theories:

1. Moral theories generate verdicts.
2. Verdicts get tested.
3. Once a verdict is confirmed, it takes up the role of evidence in favor of a moral theory.
4. If competing moral theories generate the same verdicts, and they are equally confirmed, they are deontically equivalent.

Using confirmed verdicts of moral theories as the *particular* verdicts Baumann refers to seems like a natural way to move forward, given the sustained analogy between moral and scientific theorizing. However, unlike scientific theorizing where predictions are tested against experimental results, for instance, the question of against what verdicts are to be tested looms over moral theorizing. How can we justifiably believe a verdict is true or not? We would need to ‘test’ it against the readily available non-theory driven moral evidence we have at our disposal. The list of kinds of moral evidence is long; moral testimony, judgements of common-sense morality and intuitions are a few we can cite. Arguably, we cannot just pick out one kind of evidence to our liking. All three kinds of evidence stated are likely to play some evidential role at some point. I find myself favoring intuitions, at least insofar as a primacy is to be assigned to them over the others is concerned. The following discussion does not rest on the differences between intuitions and, say, moral testimony or common-sense morality. For the convenience of not stating all the three kinds of evidence all the time, let us pick intuitions as one kind of moral evidence which we will employ.²⁰ This is also convenient to discuss Parfit, for whom we intuitive beliefs about rightness of acts (Parfit, 2011, p.415). Everything I say next will, I believe, can be applicable to other kinds of evidence too.

²⁰ See Audi (2015) for a detailed overview of intuition in ethics.

If evidence plays a similar role to experimental results in science, verdicts of a theory are to be tested against our intuition. This means that if a verdict of a theory is aligned with our intuition, then we are likely to count that *particular* verdict as evidence of the theory's truth, and the set of all such verdicts is the 'particular verdicts' Baumann proposes to assign the role of evidence when assessing moral theories.

5.2: Underdetermination of TT and NCC''

We have clarified what constitutes the 'particular' verdicts we are to assign the role of evidence in ethics. They are the verdicts that are to be tested against moral evidence, that when they are confirmed, they themselves assume the role of moral evidence too. For practical purposes, we choose intuitions as one kind of evidence to use; though I claim the following discussion is conserved its validity with different kinds of moral evidence too. Let us now turn our focus to whether TT and NCC'' are indeed underdetermined by evidence.

I have expressed my skepticism about whether TT is a genuine moral theory. If TT is not even a theory, it evades underdetermination. Baumann himself is not committed to whether TT is true or false, but rather he assumes it to be true for argument's sake, in order to show Parfit fails to show that CC dissolves disagreement among experts. To go through Baumann's argument, and illustrate how NCC'' is a better option compared to TT, I too for argument's sake assume it is a genuine moral theory.

Applying P1: 'If two moral theories (MT) can account for exactly the same evidence, it is equally reasonable to believe either of them', and P2: 'If it is equally reasonable to believe either of two

MT, we have no reason to attribute truth to one but not the other' to both TT and NCC'' will yield, assuming Kantianism, contractualism, and consequentialism are competing theories, that they are indeed both underdetermined. If competing theories are underdetermined, we have no reason to believe any one of them is true. Given that the verdicts of all 5 theories are the same, they will all be equally confirmed by our intuitions. However, as Baumann himself is quick to note, to claim that one kind of evidence exhausts moral evidence is perhaps too narrow, and that there is room for additional evidence. I will not delve into whether inclusion of additional evidence like moral testimony, common-sense morality, or theoretical virtues (Keas, 2017) can undermine P1. Rather, I want to focus on the differences between explanatory contents of TT and NCC'', and whether we can undermine underdetermination there. Put differently, we will be aiming for P3: 'If two MT contain incompatible propositions, they cannot both be true.'

When it comes to TT, P3 is tough to argue against. As we have covered in section 4, Parfit's move to combine competing theories, rather than argue for a narrower convergence, implies that TT itself contains incompatible proposition with its rivals. This is assuming, as I have argued, that there is a very strong sense in which the grounding claim of Kantianism and consequentialism, or contractualism, cannot be true at the same time. Similarly, in metaphorical terms, as I have stated previously, one cannot take three routes at the same time. Explanatory differences in moral theorizing, the routes taken atop the mountain matter, unless one is content with black-box moral theories.²¹ Arguably, this shouldn't come as a surprise, given how demanding, upon reflection, TT is. While previously, competing moral theories could have broken the evidential tie due to one explanation being more likely, or intuitive, or encompassing than other, TT removes this possibility

²¹ Here, I am referring to explanation (2) kind of explanatory claims which concern explanations as to why a moral principle seems to be true, i.e., what grounds a seemingly true moral principle.

since it reduces all explanatory claims to explanation (1) kind of claims, about which there is a convergence among TT and its rivals.

What about NCC''? Is it also underdetermined? Is the project to reduce disagreement and find some convergence among competing moral theories doomed thanks to Baumann? The answer, I'm happy to say, is a no. NCC'' is a weaker version of TT which, following Parfit, aims to argue that the argument from disagreement is not as pressing as one would initially think. It follows from NCC'' that there is considerable agreement among experts when it comes to what we ought to do, and what principles we can use to find out what we ought to do. When it comes to questions of explanation (2), NCC'' leaves the door open. It could be that optimificity is what grounds what is currently shared among experts, or it could be the property of not being reasonably rejectable. It is not a theory like TT; it is a claim about the similarity, or the extent of convergence, of our best moral theories. If there is progress in moral theorizing, we can expect the disagreement in explanatory contents to reduce more and more; thereby further reducing the force of argument from disagreement.

One might appeal to the claim that NCC'' is not underdetermined by claiming that if one of the competing theories were to triumph over the others, so to speak, NCC'' would be underdetermined. The fact that it is less committal than TT only gives it the appearance of not being underdetermined; but it could, in principle, be underdetermined in the future. To them, we can say no. Unlike TT, NCC'' is not a claim about a near-completely true moral theory. It is a claim about convergence. If, for example, Kantianism were to be shown to be *the one* true moral theory, Kantianism would still be the one true moral theory, for NCC'' is entailed by it. NCC''

makes no explanation (2) commitments, and any theory we could reasonably show to be true will involve explanation (2) commitments. NCC'' will simply be a subset of it.

Baumann (2018) asks us whether we can be happy with deontic equivalence, and leave the disagreements about explanation (2) to the side, like scientific anti-realists do. The answer is a no. We could interpret NCC'' as a statement of the progress of our moral theorizing. We can appeal to shape the future of our ethical theorizing discourse moving forward, and focus on the remaining disagreement we have. Will they ever be solved? I doubt it, but it is a criterion by which we can effectively distribute our resources. Contra Baumann, I believe the explanatory disagreement concerns not only the moral realists, but all moral theorists. Even if one is a radical subjectivist, we can justifiably demand 'Why?'

Chapter 6. Conclusion

No matter how one assesses Parfit's project, the effort is surely monumental. Throughout this work, we have delved into corners of philosophy Parfit himself probably never considered. This is arguably the mark of a good philosophical work for it gives us more and more to think about. Throughout this work of interpreting Parfit, I operated with Parfit's aim in mind: to resolve the considerable disagreement among experts. This, I believe, was done by Parfit himself. Alas, there were distinctions along the way that were missed by him, and we stepped in. Thanks to the pioneering work of Dietrich and List, and Baumann, this work allowed us to build a bridge between moral and scientific theorizing. Although I have no intention or desire to scientize philosophical thinking, I believe we found plenty of useful distinctions about structure of moral theories, moral evidence, and more.

At the end of the day, this work argues that Parfit did succeed in his aim, though some interpretive work was needed. At points, we had to acknowledge the mistakes Parfit made. Convergence works not in the exact way Parfit envisaged, or to a lesser extent. Nonetheless, anyone who believes their work to be threatened by the argument from disagreement can, for now, take a deep breath.

Finally, we have seen the force of a metaphor once again, which perhaps can remind us why Plato used them so often. To this point, I have not encountered upon a work which delved into interpreting the mountain metaphor. Interpreting it proved to be invaluable for this work. The excellent philosopher Parfit is, there was something about metaphor that leads us to think beyond what is written near 500 pages of *On What Matters*. It allowed us to delve deeper into parts of Parfit's work that were perhaps not as clear or informative as one might have hoped. This work, without

a doubt, would not be the same without it. I believe more interesting results are to be found in interpreting the metaphor.

In this work, I have interpreted Parfit's elusive mountain metaphor, which symbolizes competing moral theories converging. I have asked and answered the natural follow-up question 'What exactly is convergence?' I have argued for a narrow convergence thesis, which is to be found in Parfit's work, but itself is not explicitly argued for by Parfit. The narrow convergence thesis, or NCC'' throughout this work, obtains; hence, the argument from disagreement is resolved as Parfit aimed to do. Throughout the work, we have played with an analogy between moral and scientific theories, which led us to think about the structure of moral theories. Similarly, it is through the analogy between the two domains that we better understood Baumann's argument, and were able to respond to it.

The grandness of Parfit's work perhaps lies in how many different, and seemingly distant areas of philosophy were employed to interpret a 10-page segment towards the end of *On What Matters*. What intrigued me the most about this project was the potential to think about the structure of moral theories *qua* theories. As I did my best to answer some questions during this work, I leave it with many more than I started, the majority of which revolve around 'What exactly is a moral theory?'

REFERENCES

- Audi, R. (2015). Intuition and its place in ethics. *Journal of the American Philosophical Association*, 1(1), 57-77.
- Baumann, M. (2018). Parfit, convergence, and underdetermination. *Journal of Ethics and Social Philosophy*, 13(3), 191-221.
- Berker, S. (2019). The explanatory ambitions of moral principles. *Noûs* 53(4), 904-936.
- Bina, F., Bonicalzi, S. & Croce, M. (forthcoming). Epistemic authorities and skilled agents: a pluralist account of moral expertise. *Topoi*, 1-13.
- Brown, C. (2011). Consequentialize this. *Ethics*, 121(4), 749-771.
- Chappell, R. Y. (2021). *Parfit's Ethics*. Cambridge University Press.
- Chakravartty, A. (2001). The semantic or model-theoretic view of theories and scientific realism. *Synthese*, 127(3), 325-345.
- Dietrich, F., & List, C. (2017). What matters and how it matters: a choice-theoretic representation of moral theories. *Philosophical Review*, 126(4), 421-479.
- Enoch, D. (2019). How principles ground. *Oxford Studies in Metaethics*, 14. <https://philpapers.org/rec/ENOHPG>
- Fine, K. (2001). The question of realism. *Philosopher's Imprint*, 1(1), 1-30.
- Hare, R. (1993). Could Kant have been a Utilitarian? *Utilitas* 5 (1):1-16.
- Herman, B. (2011). A mismatch of methods. In Parfit, *On What Matters vol.2*.
- Keas, M. N. (2017). Systematizing the theoretical virtues. *Synthese* 1, 1-33.
- Lakatos, I. (1971). History of science and its rational reconstructions. In R. C. Buck & R. S. Cohen (Eds.), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*. D. Reidel.
- Majors, B. (2007). Moral explanation. *Philosophy Compass*, 2(1), 1-15.

McGrath, S. (2009). Moral disagreement and moral expertise. *Oxford Studies in Metaethics: Volume 4*, 87–108.

Parfit, D. (2011). *On What Matters. Volume One*. Oxford University Press.

Parfit, D. (2011b). *On What Matters. Volume Two*. Oxford University Press.

Portmore, D. W. (2023). Consequentializing agent-centered restrictions: A Kantsequentialist approach. *Analytic Philosophy* 64 (4):443-467.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.

Railton, P. (1984). Alienation, consequentialism, and the demands of morality. *Philosophy and Public Affairs* 13(2), 134-171.

Scanlon, T.M. (2011). How am I not a Kantian. In Parfit, *On What Matters vol.2*.

Scanlon, T. M. (1982). Contractualism and utilitarianism. In A. Sen & B. Williams (Eds.), *Utilitarianism and Beyond*. Cambridge University Press.

Shankar, R. (1994). *Principles Of Quantum Mechanics*. Plenum Press.

Sider, T. (2022). *The Tools Of Metaphysics And The Metaphysics Of Science*. Oxford University Press.

Singer, P., & Wells, D. (1984). *The reproduction revolution: new ways of making babies*. Oxford University Press.

Suikkanen, J. (2014). *This Is Ethics: An Introduction*. Wiley-Blackwell.

Väyrynen, P. (2008). Usable moral principles. In M. N. Lance, M. Potrč, & V. Strahovnik (Eds.), *Challenging moral particularism*. Routledge.

Wolf, S. (2011). Hiking the range. In Parfit, *On What Matters. Volume 2*. Oxford University Press