It's all you

From the self to the ideals of self-improvement and autonomy

Master's Thesis for the *History and Philosophy of Science Program*

Author

Hugo Dirk Hogenbirk

Student number

5511917

Supervisor

Prof. dr. Jan van Ophuijsen

Second reader

Prof. dr. Paul Ziche

No. of words

48.000 words

Table of Contents

Preface 4

Remarks 5

Introduction 7

Chapter 1: Kantian Philosophy and its Multiple Selves 11

Section 1.1: Kant's critical project and the Critique of Pure Reason 11

Section 1.2: Limiting reason; intuition and thought 15

Section 1.2.1: Sensibility and the Transcendental Aesthetic 16

Section 1.2.2: Knowledge, sensibility and the Dialectic 21

Section 1.3: The transcendental analytic: The deduction and the self 24

Section 1.4: Mental self? Bodily self? 28

Section 1.5: The empirical and transcendental self in Kantian ethics 30

Section 1.5.1: Freedom and determinism 31

Section 1.5.2: Transcendental and empirical action 32

Section 1.5.3: Giving the law to oneself 35

Section 1.6: Conclusion regarding the relation between empirical and transcendental self 38

Chapter 2: Normativity, Virtues & the Self 41

Section 2.1: Virtue ethics and Kantian ethics 41

Section 2.2: Norms and reason 46

Section 2.3: A problem of normativity for virtue and its (dis)solution in the dual-self 51

Section 2.4: What is Kantian virtue? Does Kantian virtue do what it should? 53

Section 2.4.1: Kant on virtues in the Groundworks and the 2nd Critique 54

Section 2.4.2: Kant on virtues in Metaphysics of Morals and Religion within the Bounds of Reason 57

Section 2.5: Striving 60

Chapter 3: Neo-Kantians on the Self 62

Section 3.1: Korsgaard and constitutive principles and defective action 62
Section 3.2: Determinedly evil and the capacity to fail 64
Section 3.3: Freedom and evil 68
Section 3.4: Intermezzo through Schelling; symmetry and asymmetry 69
Section 3.5: Kant: between *Wille* and *Willkür* 72
Section 3.6: O'Neill: constructing reason and prying apart persons, freedom and actions 79
Section 3.7: O'Neill's proposal; self-legislation 83
Chapter 4: Extensions of the Self 87
Section 4.1: Aiming for perfection 87
Section 4.2: Ascetics 90
Section 4.3: Transformative power of knowledge of the self 93
Section 4.4: Narrative understanding of the self and practical identity 98
Section 4.5: Moral Education 103
Conclusion 108
Bibliography 111

Preface

Thank you to all who have helped me throughout this long process of getting my thoughts in order and eventually, on paper. Without you all, this would have been a severely diminished product. I take great pride in the result that now lies before you (in a form, again, only possible due to your help). It is a thesis that starts with technical analysis of the philosophical system of Immanuel Kant and slowly but steadily 'devolves' into the question of how we should aim at self-improvement. I wanted to combine these two things, for they both interest me and they show the possibility of combining the *abstract* with the *concrete* in one philosophical project. The topic of self-improvement (undertaken by ourselves) at some point started to fascinate me; I aim to better myself and others aim to better themselves. But, does this not mean we are unhappy with who we are? How can one and the same thing dislike its entire being and aim to change it in one fell swoop without falling into paradox? The answer came from Kant's proposal to split up the self, to differentiate between our contingent existence and our intellectual existence. We improve by allowing the one to be dissatisfied with, and act upon, the other. This is one of the two intended meanings of the title. All these different facets of personhood we will be discussing in the thesis, *they are all you*. You incorporate all of them, and to be human is exactly to share in all these facets. However, *It's all you*, also takes on a call for autonomous action, for, *you yourself are also up to you*. Self-improvement and autonomy go hand in hand; taking the construction of your own personhood into your own hands is aiding yourself in becoming an autonomous creature and becoming autonomous aids in your project of self-construction. We should aim at self-improvement in the most banal ways, but we should also make sense of *developing our capacity for self-improvement*, which is, in a broad sense, in accordance with Kant's idea of autonomy being characterized by (being capable of) giving the law to oneself. Connecting the insight that our own self is split-up and that we should attempt to better each of these parts; I hope some of you will find this connection productive in *philosophical reflection* or in a *concrete project of self-improvement*. But most of all, I hope you enjoy the read.

Remarks

Chapter 1: In the first chapter (and sometimes throughout the thesis) we make use of a number of ways to refer to those creatures which can be characterized as being self-reflective and receptive of the outside world. Self, subject and person are all used and they are not clearly delineated. They do however, put a different emphasis on the being that is under discussion and tease out different elements. *Person* is used from the third-person perspective while *self* is used from the first-person. When I want to say that there are many creatures of the relevant sort, I might say there are many *persons*, who come across as being this or that way. This is also the broadest, most all-purpose way of referring to humans in their capacity as reasonable beings. *Self* I would use when discussing the ways a person relates to him or herself. This points our attention to the reflective side of a person's existence. A self is always *someone's self*. Subject is most often used when it is juxtaposed with the word 'object' (the subject vs. the object) and sometimes aims to get attention for the *receptive* side of a person's existence. In chapter one we eventually come to new terminology (empirical self and transcendental self) which captures the distinction between the reflective and the receptive and given, parts of a person's existence more neatly. These terms are used as truly technical vocabulary, the aforementioned are used in their everyday meaning and applied also on the basis of what is stylistically nice in one or another sentence. On the question of whether this is not a departure from how Kant uses these words, we like to invoke a summary section of Longuenesse in her recent work *I, me, mine*.[1] The section is called *Ambiguities in Kant's Terminology*. Here she discusses a number of terms in the Kantian corpus, including *subject* and *self*. Person is not included and does not seem to be a technical terms at all. Self is argued to be a very ambiguous term for Kant without clear unified meaning. Subject does have more specific meanings Longuenesse claims, but this one is also ambiguous. Sometimes referring to what we will come to call a *transcendental self*, the person in it's mere existence as a reflective creature. But sometimes it is also used to refer to soul, which is quite close, though distinct, of what we will call the *mental self*. We believe it is justified given that these terms are not given strict technical meaning in the Kantian system to use the words as style and association dictate.

Chapter 2: In chapter two and from chapter two onward we make use of the term *virtue ethics*. There is a related term called *virtue theory* which we will not use. There is an insistence from virtue ethicists to

[1] Longuenesse, I, me, mine, p.103-112

mark the difference in the following way: virtue ethics is when one does moral philosophy that in some way, shape or form, takes virtue as the *primary* moral category or unit and virtue theory is when other moral philosophers have accounts of virtue in their moral theories without making virtue morally primary.[2] Accordingly, what we attempt in chapter two should be called *virtue theory* not *virtue ethics*. We do call it virtue ethics for two reasons. Firstly, we disagree that it is useful to drive a wedge between these two ways of accounting for virtue, especially since a lot of virtue ethics is motivated by *the lackluster capacity of other philosophies to account for virtues in their theories*. However, more importantly, we believe that the names themselves are not apt. A Kantian account of virtue is not *merely* a Kantian who is theorizing on virtue. What we attempt in chapter two is to make sense of virtue *as an ethical object*. Just because the primary moral force of the theory does not derive from virtue, this does not make it any less an ethical account of virtue; the attempt is to make virtue *a truly moral category within the Kantian framework*. For these reasons we have opted to hold to the term virtue ethics and disregard virtue theory in the main text.

Chapter 3: The first two sections of this chapter are slightly adjusted sections from a paper I have written before called *Failing to Reason*. The reason is that there, I discovered some issues with the neo-Kantian insistence on deflating the self in Korsgaard's work. That found its way into a more general discussion in chapter three on what Kant incorporates that makes him impervious to that criticism, and how O'Neill attempts to deal with the problem. Yet, the text on Korsgaard was already adequate and for that reason adjusted to the thesis.

General remarks citations and references: In citations we have opted to show our own emphasis by printing the text bold. The works of Kant are referenced by use of the numbering of the *Akademie* edition of his works, except for the first *Critique* which is referenced by it's own standard numbering. We have also added page numbering for these works which reference the translations that have been used for the citations. In citations we used square brackets to either enter explanatory words that are needed to compensate for a loss of context that would make the citation hard to understand, or we used them with three dots between them to indicate that we have left out a part of the text in between two cited parts, or the start or end of a sentence.

[2] Hursthouse and Pettigrove, Virtue Ethics

Introduction

It is quite common in philosophical circles to assert that in the 20th century we have made a 'turn' or 'shift'. For example the so called 'linguistic turn', where, especially in the analytical/Anglo-Saxon tradition, we find an identification of language as the place where philosophical inquiry should be focused. At times it even signifies the conviction that a lot of (or all) classical philosophical 'problems' are little more than 'confusions' to be solved by analysis of the language they are expressed in, which in the end leads to that, as Wittgenstein puts it: *"the result of philosophy is not a number of 'philosophical propositions', but to make propositions clear"*.[3] In continental philosophy we may discern similar movements. Not only do we also find there a linguistic turn in de Saussure, Foucault, Derrida and so on, but starting with Marx, we find an explicit shift of attention towards the concrete situation of the person for cultural formation and belief formation. The importance of this for philosophy is something Marx quite famously expressed in the statement: *"The philosophers have only interpreted the world in various ways; the point, however, is to change it."*[4] From Marx onward we find an appreciation for the way background, class, sensory apparatus and other empirical factors, function in the production of human action and thought, through for example Freudian psychoanalysis, Foucauldian discourse analysis and a Heideggerian focus on the *alltagliche* (everyday) dimension of self-aware beings. And yes, again we find that on the other side of the Channel, in analytical philosophy, not only a linguistic turn, but also a shift towards the concrete person has taken place. In the work of the historian of science Thomas Kuhn, we find a picture of science as produced by, in the end, human factors, which do not *guarantee* a rational nor cumulative picture of science. In the naturalized epistemology of Quine, we find a wholesale acceptance of the problematic nature of 'philosophical intuition' on the basis of the fact that philosophers stand in concrete situations themselves, which in the end leads to the denial of the possibility of classical epistemology.

The goal of the previous section is to illustrate the following point: in recent times philosophy became more *concrete,* all the while becoming less interested in the *abstract*. That is to say, it has put greater emphasis on the importance of the concrete, or has embraced some form of the concrete wholesale, denying the possibility of the abstract. Whether it is the exchange of the (abstract) propositions for the (concrete) sentences, or the exchange of an (abstract) Hegelian view of history as an unfolding of

[3] Wittgenstein, Tractatus Logico-Philosophicus, prop 4.112/p.77
[4] Marx, Theses on Feuerbach, p.15

the concept of freedom for the (concrete) Marxist view of history as an unfolding of economical relationships, there is in these changes a move towards the concrete in general.

There are good reasons for a move toward the concrete. The reason posed by Marx, *that philosophy, by remaining abstract, lost its applicability to the world*, is one of them.[5] Another unproblematic reason is that 'variety is the spice of life', why should we *not* investigate more concrete phenomena? However, this is often not the actual reason stated by the philosopher who would investigate the more concrete. Most often heard is the position that the abstract is (now) *untenable*. It is when philosophers hold this view that the concrete and the abstract become opposed, i.e. mutually exclusive. For example, one can mean one of two things when stating that we should investigate sentences as truth-bearers. It can mean that we have always overlooked sentences as truth-bearers, besides the already investigated propositions. In this case, the two can co-exist. But it can also mean that we should investigate sentences as truth-bearers *instead of* propositions, or even more stringent, that it is precisely because we understand propositions to be untenable that we now should investigate sentences. In these two last situations, the abstract and the concrete have become mutually exclusive.

This master's thesis will occupy itself with a specific instantiation of a topic that has in the 19th and 20th century been approached more and more concretely; the self. As we pointed out, the 20th century provides us with many major philosophers who investigate the concrete way that the subject is created and can be viewed as a material object, with authors like Heidegger, Foucault and Ryle at the forefront of this approach. What this master's thesis will investigate is the *interplay* between more abstract conceptions of the subject and the more concrete. Not historically, how the one idea led to the other, but systematically, *how one's abstract self and one's empirical self, interact*.

We will argue for compatibilism between concrete and abstract conceptions of the self. We will argue that, to make philosophy effective, to make it applicable by anyone interested in getting to know

---

[5] A qualification is in order. Russell, in a 1960 interview with Woodrow Wyatt, explains Marx' statement by describing him as "there are some philosophers who […] exist to upset it [the status quo]" and takes the contrary position that the task of the philosopher is to understand, not to change. This is not how we understand the remark by Marx. Marx' reaction to Hegelian dialectics was not just an existential *call to action* for the philosopher as a person, but also an identification of a problem with the abstract nature of the philosophy; namely that one could not even hold out hope for *any* applicability to life or the world with this kind of philosophy. It is in this sense that we would agree with Marx (if he intended it so); we should *wish*, or *hold out hope for* philosophy to influence the world. If we follow Russell's interpretation as the statement being inherently *revolutionary*, we would rephrase the sentence as: It is the philosophers job to *hope* to change the world *through* understanding.

(and to improve) herself, a one-sided philosophical story is *devoid of power*. We will draw on the works of Kant, most importantly the *Critique of Pure Reason*, where he introduces the subdivision of the self into two parts. The first, abstract, self is what Kant calls the transcendental self, the second, concrete, self is called the empirical self, further divided into an outer, bodily self, and an inner, mental self. This will be our model for our compatibilism. Chapter one will be concerned with the explication of this Kantian model of the *whole person*.

But of course, we are not only claiming that it is theoretically consistent to make sense of *both* the concrete particular existence of the person while maintaining that it also contains an *intellectual, reason-directed* side. We will argue in chapter two that this account of the self can be made productive regarding the issue of *virtue*. Virtue ethics often takes a combative stance toward Kantian ethics, and a number of the arguments against the Kantian scheme are due to the Kantian's inability to account for a number of moral phenomena. Most importantly, the Kantian is argued to be incapable of making sense of *virtue*, dispositions towards doing the good, as its own moral category. We will argue however that the Kantian scheme, especially with a focus on the dual structure of the self, as we will introduce it in chapter one, is *very potent in making sense of virtue as a moral category, more so than pure virtue ethics*. It is our specific conception of personhood that will allow for this.

There are however Kantians who would take another route to make sense of the person. These are often less metaphysically involved than the Kantian picture we will sketch for you throughout the thesis. Christine Korsgaard understand the self as essentially *self-constituting*, whereas for Kant, there is a substantial self (not merely in the physical world, but also an *intellectual* transcendental self) that exists before any self-constitutive action occurs. Onora O'Neill tries to divert the attention away from *any* account of the self, arguing that we should investigate what is reasonable to do and *distribute reasonableness* or *construct it*, based on genuine appraisals of the people around us (again, for Kant, there is a matter of fact about personhood, before the act of a *construction of another's reasonableness*). We will discuss the problems for both these views and argue for the merits of the Kantian picture we have presented, through a discussion of *the problem of being evil* and *the problem of giving the law to yourself.*

However, as is stated in the subtitle of the thesis, there is throughout the thesis a theme of *self-improvement*. This has mostly found its expression in chapter two, and the final chapter four. In chapter two, we have argued for the importance of breeding a virtuous character and made clear from where the normative force for doing so would derive on a Kantian picture. In chapter four, we will make these insights more concrete by discussing techniques for self-improvement that are related to the developed

model of the dual-self. The hope is that the thesis thus becomes a coherent whole, from abstract conceptualization of the self, to the normativity of virtue and self-improvement to eventually, concrete techniques of self-improvement that have the split person as their model.

Chapter 1: Kantian Philosophy and its Multiple Selves

The self-identical transcendental self and the contingent ever-changing empirical self

Abstract: In this chapter we will look at the work of the philosopher Immanuel Kant. The goal of this is to introduce and analyze his dual concepts *transcendental self* & *empirical self*. For this, we first introduce Kant's *critical project*, a philosophical program spread out over Kant's writings from 1781 (the publication of *The Critique of Pure Reason*) onward. We will see that the Kantian conception of the self is at the center of his critical endeavors. This finds its expression in Kant's distinction between the *formal and general* aspects of being a self, which we dub the transcendental self, and the *concrete and particular* aspects of being a self, which we dub the empirical self. The empirical self we find is again split up, one part observable by other beings and spatiotemporal, the bodily self, and the other private and merely temporal (non-spatial), the mental self. We introduce these concepts through analysis of their importance for Kant's work on *theoretical reason*, (the transcendental and empirical self in the context of being *thinking and knowing beings*). However, we will find that Kant elaborates on the properties of the self in the context of ourselves as *acting beings* (practical reason). Here we find that the transcendental self must be understood as free, that action is specific to us split beings and that giving the law to oneself implies the possibility of understanding oneself as both active binder and passively bounded.

Section 1.1: Kant's critical project and the Critique of Pure Reason

In the *Preface* to the B-edition (the second, revised, edition) of his monumental *Critique of Pure Reason*,[6] Kant described his critical philosophy as a 'Copernican Turn', a phrase about which many words have been spilled, both to clarify what Kant meant and whether it even is a sensible analogy for this project. On one reading, just as Copernicus put the sun central in the universe (which we now understand to be a claim about the solar system), Kant has put man central.[7] Here, the phrase 'Copernican Revolution' points to taking something new as the central element of one's thinking; we change what revolves around what. It also zooms in on the *result* of Copernicus' work. A more elaborate reading[8] however focuses on the *methodology* of Copernicus. In this case, we draw an analogy between Copernicus' insight that when

[6] Kant, Critique of Pure Reason, B XVI – XVII/p.110
[7] In this same vein, an old professor of mine once claimed that the Copernican analogy made little sense; Copernicus put man (the earth) less central to our cosmology, while Kant had put man completely central to our thinking. Kant should be understood as an *anti-Copernican* on his view.
[8] See *Kant's Copernican Analogy* by Smith for this view, which we paraphrase here.

accounting for motions, we can both ascribe motion to the viewer and the viewed[9] and Copernicus' choice to ascribe it to the viewer (the earth) instead of the viewed (the stars)[10], and Kant's insight that when accounting for knowledge in general, we can ascribe parts of its origin to the subject, as Kant chooses to do and argues for, throughout his critical project.

Despite unclarity about the precise way we should interpret Kant's remark, what is common in most readings is that Kant puts forward the *self* as a (genuine) *source of our knowledge*. Of course, the way we 'are' productive of knowledge is not in the same, concrete, way as we find in Copernicus.[11]

What Kant is after in his critical project (especially *in Critique of Pure Reason*) is to chart the limits of reason; to show us where we can expect reason to generate knowledge, and where we cannot expect reason to deliver us knowledge.[12] Although that is the goal, the main innovation Kant proposes to achieve his goal, is to put the requirements for knowledge being knowledge into the subject, leading to what Kant calls the 'transcendental' character of his philosophy:

> *I call all cognition transcendental that is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible a-priori.*[13]

By turning his investigation to the subject (and not the objects), and analyzing what the prerequisites are for having knowledge at all, Kant tries to find a way to show that reason is more capable than some earlier philosophers would assert (he famously positions his work against that of the empiricist and sceptic

---

[9] If this is not immediately clear imagine two boats out on open sea. When the two ships are not moving relative to each other, there is no way to differentiate by looking at the other ship whether you are both moving, or are both lying still in the water. In both cases, the motion of the other ship would seem to be that of a ship that is not moving. In a similar fashion, if the other ship appears to be moving to the left (that is to say, relative to you), there is no way, by reference only to the other ship's movement, to differentiate between you moving to the right, or the other ship moving to the left, or you both moving to the left where the other ship is just moving a bit faster than you. So through this we find in Copernicus the possibility of attributing appearances not only to the things viewed (the other ship), but also to the things that view (our own ship).

[10] Copernicus, On the Revolutions of the Heavenly Spheres, p.40

[11] Although we will see that the implicit limiting of what can be known through the concept of relative motion bears a lot of resemblance to how Kant limits thought. We provide a comparison of modern philosophy of physics and Kant in footnote 38

[12] Kant, Critique of Pure Reason, A XII/p.101

[13] Kant, Critique of Pure Reason, B 25/p.149

philosopher, David Hume).[14] He tries this by showing how some elements of knowledge can be presupposed in the subject itself. At the same time, having gotten a clear view of the workings of the subject and the properties of knowledge, Kant can still limit reason to its proper domain, resulting in his famous rejection of the cogency of the claims to knowledge of previous metaphysicians, which are characterized as capable of merely the "…art […] of metaphysical jugglery" and "…sophistical tricks" .[15]

All of the above is a commonplace reading of what Kant sets out to do in his critical project. However, for what follows in this chapter, this is not always the case. So, before we continue onward, we should make an important qualification regarding the relative status of our reading and rendition of Kant compared to more general Kant scholarship. There are multiple ways of interpreting the vast and sprawling *critical system* of Kant. This means that inevitably we will be presenting you with a specific interpretation. The most distinct way in which our reading differs from many modern readings is that we believe Kant *hypostasizes* the self some of its facets. That is to say, we take Kant to believe the self, even in its *supra-empirical* existence, to be a proper *thing*. There are both reasons to discard and to hold to this reading. Textually, in 1.5.2 and 1.3, we find citations of Kant that seem to point in this direction, especially Kant's views on autonomy in his moral philosophy underpin our claim. In 3.6 we find a reading (though not defense) of Kant that is similar, provided by Onora O'Neill.[16] However, in his famous *paralogisms*, Kant discards many substantive claims previous philosophers have made about the self. This is reason for readers of Kant to discard the substantive view of the self (Longuenesse in *I, me, mine* takes such a general approach). Another important place to point to is *The Canon of Reason* in the *Critique of Pure Reason*[17]*,* where Kant takes a less substantive attitude towards what is implied by the possibility of moral philosophy. Throughout his moral works however, Kant seems to rebound from the lean conception of morality we find there, and we find ourselves sin the company of Henry Allison who tells us in his book *Kant's theory of freedom*:

> *[…] it is not surprising that that his discussion of morality in this work [the first Critique] is extremely sketchy.*[18]

---

[14] Kant, Prolegomena to any Future Metaphysics, 4:257/p.7
[15] Kant, Critique of Pure Reason, B88/p.200
[16] O'Neill, Constructions of Reason: Exploration of Kant's Practical Philosophy, p.i
[17] Kant, Critique of Pure Reason, B823-B859/p.672-690
[18] Allison, Kant's theory of freedom, p.66

The *Canon of Reason,* is for others, like Resscher in *Kant and the Reach of Reason* (see for example p.172-174), an important part of Kantian text for their 'lean' reading. We however believe Kant leaves (and must leave) room for some substantive self. This is the reading of Kant we will be expounding in the rest of the chapter. In chapter three however, we will explicitly argue for the merits of this reading by taking the substantive Kantian account and pointing out its merits against the neo-Kantians Christine Korsgaard and Onora O'Neill, who take a more deflationary stance. With this, let us continue to expound the Kantian system.

The Kantian project is heavily based on the concept of a subject, or a self, which itself is asked to bring bits and pieces to the table in the construction of knowledge. However, this is not an affirmation of vulgar relativism, where due to the knowledge being produced by the subject, it becomes knowledge only valid relative to the subject. It becomes knowledge relative to the idea of *any knower whatsoever*, but not relative to any specific knower. In that sense we must understand Kant's remark that that which is transcendental is only that which occupies itself with the possibilities of *a-priori* modes of cognition.[19]

There are situations where things happen that point in the direction of more concrete forms of influencing one's own perceptions; some people's eyes might identify an object as being green, where the other, being severely colorblind, sees only a somewhat grey object. However, this is not the sort of 'dependence' on the subject that Kant is thinking of. For Kant's intended 'dependence on the subject', Kant needs a conception of the self that is *as narrow as it can be*, that is to say, that incorporates no more in the subject than exactly that what is *required for it to be a subject*. As we will discuss further on, for Kant the bare *stipulative* minimum for being a subject is to be able to *think the 'I think'*. From this *synthetic unity of apperception*, Kant derives a broader account of the *minimal subject*, namely the *minimal empirical subject* incorporating *categories of the understanding* and the *forms of intuition*.[20] This conceptualization helps Kant, because if he can prove that even within the most narrow account of a subject we find the prerequisites for some sort of knowledge, we can rest assured that *we* can in principle

[19] A-priori knowledge is independent of specific ways the world is or we ourselves are, and it can be ascertained without recourse to empirical content.
[20] How this hangs together exactly, and what supplementary assumptions are required by Kant will be discussed when we arrive in 1.3

be justified in this sort of knowledge.[21] That is to say, if we have the pre-requisites for having knowledge of cause and effect within this minimal conception of the self, we know that it is *possible* (given the right empirical content to back this knowledge) to have knowledge of concrete instances of cause and effect. Likewise, once Kant shows that for a subject, as subject, knowledge of certain types of things is impossible, we can rest assured[22] that we certainly know we cannot know anything of these things.

Kant needs his narrow self, for his project to succeed. We will call this conception of the self the *transcendental self*. This is to indicate the importance of this self for the specific transcendental nature of his project.[23] If we want to attain knowledge by laying parts of knowledge in the subject, this should not depend on the specifics of one or the other person for the knowledge to be proper knowledge, so to delimit reason we require a concept of the subject as subject: the transcendental self. The rest of our self, the contingent part that is not pinned down by my being a knower at all, is our empirical self. In the coming sections we will delve deeper into this pair of concepts through a further rendition of Kant's philosophy.

Section 1.2: Limiting reason; intuition and thought

The Critique of Pure Reason can be split up in many different, insightful, ways. What we will start with is the division Kant draws between part one and part two of the *Doctrine of Elements* (which make up the bulk of the *Critique*). This division is that between the *Transcendental Aesthetic* and the *Transcendental Logic*, which roughly correspond to a transcendental discussion of *sensibility*[24] and a transcendental

---

[21] As Whitehead puts it beautifully, we are in this case justified in applying the because we will never: "[…] *catch the actual world taking a holiday from their sway"*. (Whitehead, Process and Reality, p.4)

[22] The reason this should re-assure us, and not make us feel uncomfortable at all, is because Kant has a plan for these sorts of objects. For, when something is completely exempt from being an object of knowledge, we can be sure that we will also never find a way to *disprove* the existence of that object. It is this feature Kant will build on to formulate such objects as the "Postulates of Practical Reason", where these object can still figure in ethics *as if* they exist.

[23] We could also have called it a transcendent self. Both names are apt. The transcendental self is epistemologically important because it allows for the transcendental project itself; it allows us to understand the idea of a narrow self that we can *count on*, to aid us in our acquisition of knowledge. However, we could also call it transcendent if we would describe this self metaphysically, for it is (we will see this later more explicitly) the noumenal side of the self, a self that exists outside of the empirical world.

[24] The reason we put sensibility down as a noun and thinking as a verb is because one of the most important division between the two is that the one is *given* while for the other, we are ourselves, actively, responsible. However, putting

discussion of *thinking*. The discussion on thinking is further subdivided, into the *Transcendental Analytic* and the *Transcendental Dialectic*, which roughly correspond to a discussion about how (correct) thinking is possible and a discussion about how it is possible and probable that we judge in way that cannot be justified because it transcends the limits of empirical knowledge.

In this section we will discuss the *Transcendental Aesthetic* and the *Dialectic.* Understanding these parts allows us to understand the *destructive* character of Kant's work, which is a good point of departure for understanding the more complicated *constructive* parts of Kant's work in the Analytics and his ethics.

Section 1.2.1: Sensibility and the Transcendental Aesthetic

The starting point of Kant's critical philosophy is an analysis of sensibility, insofar as it is sensibility. That is to say, we are not looking at specific variants of sensibility, but at what it means for sensibility to even be sensibility. For example, color cannot be expected to figure in an analysis of sensibility as sensibility, as color is a property tied to a specific type of our sensible existence (sight).

Connected to sensibility are, for Kant, the two particulars of sensibility, namely intuition and perception. Intuition is what for the Humean would be an impression[25]; although I'm not aware of an intuition/impression (it might occur in my peripheral vision), sensibility is providing me with colors and shapes. A perception is an intuition of which I am aware that it is given to me.[26] I am right now focusing on words in front of me, I am aware that these words are given to me. In this case, I have a perception of the words. If however, there were words in my peripheral vision, without proper attention, I have an intuition of the words.

For Kant, one of the defining features of sensibility is its given nature. Where we can on Kant's view always decide to judge issues according to our will, this freedom does not exist in the case of

---

down sensing is also possible, for it is indeed 'something we do' in the sense that *we* add the forms of intuition space and time (although we cannot fail to do so in any instance). Kant puts it like this: "[…] *there are two stems of human cognition […] namely sensibility and understanding, through the first of which objects are given to us, through the second of which they are thought*" (Kant, Critique of Pure Reason, B29/p.152).

[25] Hume, An Enquiry Concerning Human Understanding, p.9

[26] Kant, Critique of Pure Reason, B161/p.261-262

sensibility. We are of course capable of influencing what intuitions and perceptions we might come upon; when looking at something unpleasant I might turn away my head, but I cannot decide not to see the unpleasant thing by sheer force of will. The unpleasant perception is given to me, not up to me. The contemporary philosopher Rödl makes the point as follows:

> *His [Kant's] thought may be put as follows. An act of knowledge is an act of thinking an object and as such bears the form of thought. But as the object known is not given in this act of thinking, but in a distinct act of sensibility, the form of knowledge is not simply the form of thought; it also reflects the form of our sensibility. And these differ: thought is spontaneous, and that is, infinite; sensibility is receptive, and that is, finite.*[27]

This receptivity is a core aspect of perceptions and intuitions, they are *given* to us, spontaneity is a core aspect of thought, we actively think. We are limited in the case of perception by the limits of what can be intuited at all, i.e. we cannot sense the supra-empirical and thus sensibility is finite. In thought this limit does not exist, we can easily think, or conceptualize, supra-empirical objects, and thus thought is infinite, or, unbounded. There is a second sense however, that we would add to Rödl's, in which we are limited in the case of sensibility. For we can only perceive *what is actually out there to be perceived*. I cannot look out of my window and *not* see a tree. The world limits my sensibility; from this limitation of sensibility we also gain its justificatory strength. Sensibility, being limited by how the world is, stands in contact with the world. Thought on the other hand does not share this connection, we can conceptualize fictional objects, that could be out there, but *de* facto, are not. Sensibility is again limited and finite, thought is unbounded and infinite.

But, understanding Kant as essentially stating the view that sensibility has a given nature is a strange way of characterizing him. This idea of receptivity was already contained in the earlier works of empiricist philosophers like Berkeley and Hume, who likewise claimed that we would have to content ourselves with what was given to us by the senses. Kant's famous addition was, as we discussed, his Copernican Turn, which means we believe parts of knowledge to be internal to the subject. Kant argues for this in the case of sensibility as well. Specifically, space and time are argued to be the *forms of*

[27] Rödl, Review: Transcendental Philosophy and naturalism, p.485

intuition.[28] For a perception to even be a perception, and to be anything to us, it is required that it is captured in the forms of intuition. This also means that in every perception and every intuition, we will encounter time and space, and that everything we encounter in perceptions have some spatial and temporal ordering.

An important point to make about Kant's Copernican Turn in the analysis of sensibility is that although the innovation is that we now play a role in the formation of intuitions (we are the ones who add the forms of intuition) we are not free in adding them. In one sense, we add space and time, in another sense, intuitions are simply given to us in adherence with the forms of intuition. In one sense, yes, *we* do it but at the same time, we are not *doing* it.[29] This generates an important distinction between the *Aesthetic* and the *Analytic* (in the *Analytic* we also find that the self adds to knowledge, now in the realm of thinking) because, in a way, the proofs Kant gives in the *Aesthetic* do not actually change our thinking. The correct use of the senses does not depend on our knowledge of the functioning of the senses because we cannot by reason alone change what is given to the senses. Kant states that the function of the *Transcendental Aesthetic* is derivative, being necessary only when discussing the *categories of the understanding* (more on which further on)*:*

> *We have above traced the concepts of space and time to their sources by means of a transcendental deduction, and explained and determined their a priori objective validity.* ***Geometry nevertheless follows its secure course through strictly a priori cognitions without having to beg philosophy for any certification of the pure and lawful pedigree of its fundamental concept of space.*** *Yet the use of the concept in this science concerns only the external world of the senses, of which space is the pure form of its intuition, and in which therefore all geometrical cognition is immediately evident because it is grounded on intuition a priori, and the objects are given through the cognition itself a priori in intuition (as far as their form is concerned).* ***With the pure concepts of the understanding, however, there first arises the unavoidable need to search for the transcendental deduction not only of them but also of space,*** *for since they speak of objects not through predicates of intuition and sensibility but*

[28] Kant, Critique of Pure Reason, B36/p.157

[29] A good way of putting this difference can be found in Ferrarin who argues that we should view sensibility as *receptive* (it requires something given by a thing that is independent of it) but not as *passive* (intuition still has work to do by putting the given within the forms of sensibility, time and space). (Ferrarin, The Powers of Pure Reason, p16)

through those of pure a priori thinking, they relate to objects generally without any conditions of sensibility [..][30]

We never intuit, or percept, *wrongly*. Space and time are always in the form of intuition, and in a way, it is not up to us to make mistakes about this. We could say, that is why there was no need for a dialectics of sensibility, no need to explain our bad application of sensibility, is because there is no bad sensibility and there are no bad intuitions on its/their own right.[31] This in contrast to the *Transcendental Logic*, which occupies an analysis of thought. For we can think wrongly, and it is precisely because of this that Kant requires a dialectics of this topic, to show besides how it is possible to think correctly, to also show how it is possible to think wrongly.

Before we continue to discuss this, there is one more important distinction that Kant introduces in the *Transcendental Aesthetic*. It is the distinction between outer and inner sensibility. The most common way to understand sensibility is in its guise as outer intuition. The taste of my food, the sight of the tree outside of my window, the sound of music, these are all forms of outer sensibility. The objects that play a role in the perceptions are all both spatial and temporal. But, besides outer intuitions, Kant introduces inner intuitions as intuitions of that which lies within us. These inner intuitions are exactly those sort of intuitions *which do not make use of the form of space*. That is to say, while outer intuition is formed by space and time, inner intuition is *only formed by time.*

There are two ways to try to understand this distinction, to show how time and 'innerness' are connected. Kant's deductive method we will apply second; first let us look at examples. Say you are angry. On the one hand we can then point to all sorts of things that would belong to outer sensibility to make sense of this[32]; a red and swollen face, screaming, flashing eyes, etcetera. These are all perceptions that

[30] Kant, Critique of Pure Reason, B119-120/p.221

[31] This is both intuitive and counterintuitive. It is intuitive because clearly, one can only claim incorrect things when one *thinks about, or judges*, one's given impressions. At the same time, there are cases where the claim that we perceive 'incorrectly' seems sensible, for example when *dreaming or seeing a fata morgana*.

[32] Of course, whether we have these powers of introspection, of inner sensibility, is a disputed matter. In *The Concept of Mind*, Ryle argues that such a concept of inners sense is not tenable. However, most of his arguments fall short of being convincing. First is Ryle's contention that we have to accept that at times the information gained by introspecting might be gotten from the outer senses as well (Ryle, The Concept of Mind, p.157-158). From this he would conclude that it would not be weird to say that *all* information seemingly gained by introspection might *actually* be gained from the outer senses. However, this argument fails if one is not already predisposed to want to do away with inner sensibility, for yes, it might be that all introspective thoughts are actually, unawarely gained from

are outer and incorporate both the form of space and of time. However, these are not the only sort of perceptions that may tell me that I am angry. For others, outer perceptions are the only way to gain knowledge of whether I am angry or not. But for me, knowledge of my anger does not have to come through outer sensibility. It is more exception than rule that we say something like: "my hands are shaking, I must be angry". More often we simply perceive ourselves as being angry, without reference to our shaking hands or other sorts of outer intuition. Although difficult to further describe, the feeling of anger, *that you are angry*, is exactly an internal perception. Now we can wonder, is this indeed without the form of space? And it seems to be so; my anger thus perceived is not somewhere and there is nothing 'to the left or right' of my anger. So, anger seems to be without the form of space, it does not allow for questions of the shape or location of the object. Does anger still incorporate the form of inner intuition; time? Clearly it does, without reference to my shaking hands, I can assert that "I was completely calm until only a moment ago". So through example we see that indeed, inner sensibility consists of intuitions precisely only caught in the form of time. As Kant puts it:

> *Time is nothing other than the form of inner sense, i.e., of the intuition of our self and our inner state. For time cannot be a determination of outer appearances; it belongs neither to a shape [Gestalt] or a position, etc., but on the contrary determines the relation of representations in our inner state.*[33]

Time is the form of inner sense and space the form of outer sense. However, what we perceive outside of us is not merely caught in space, objects move and are thus also in time. This makes sense, because,

---

other sources, but it might just as well *not be so*. Also, that there is some congruence between my outer sensibility of myself and my inner sensibility of myself is exactly what we *want*. If we later identify both to be a part of the empirical self, they better well 'fit onto each other'. Ryle's second argument is that when psychology used introspection as a method for its science, it did not make much progress (Ryle, The Concept of Mind, p.158-159). This again assumes a predisposition toward a rejection as real of that which is not easily investigated 'scientifically', and of course in the end, the failure of this introspective science could just be a historical coincidence. Thirdly, Ryle argues that certain mental states would resist introspection, because they: *'cannot be coolly scrutinized'* (Ryle, The Concept of Mind, p.159). There is an illustrative misconception underlying this argument of Ryle. The Kantian scheme will help us to understand it. We do not have to coolly scrutinize to undertake introspection, because introspection, understood as inner sensibility, is not an act of thought at all. This is of course Kant's big distinction, the inner sense, like the outer sense, is distinct from thought. We might wish to *think* coolly, but we do not need to *sense* coolly, to sense coolly is a category mistake.

[33] Kant, Critique of Pure Reason, B50-51/p.163

before we can perceive objects outside of us, we need not only understand them *in their outerness* (i.e. spatiality) but also make them inner (for any perception is inner in the sense that I order all perceptions as being before or after another). Every perception is in the form of time, and those that are outer, *also*, in the form of space.

So for Kant, the difference between outer and inner perceptions, and the necessity of purely inner perception, is shown by arguing that, given that some intuition is not spatial, that it must be inner. This is done by pointing out that, to understand some intuition to not be inner to me, I need some way to conceptualize it as not being in me, and being somewhere out there. For something to be outer, more generally, we need a way to say it is here and not there. But as we argued, anger is in no specific place at all, it simply *is without location,* in stark contrast to it having a position in time. So, for something to be outer simply *is* for Kant to assert that it is spatial.

Section 1.2.2: Knowledge, sensibility and the Dialectic

Earlier we pointed to the difference between thought and intuition, thought being spontaneous while intuition is given. Both elements are according to Kant important for the creation of knowledge, as he famously put it:

> *Thoughts without content are empty, intuitions without concepts are blind.*[34]

Kant's point is somewhat intuitive, when we think without having *input* for this thinking, our thoughts have no content to work on. And clearly, without having thoughts about our input, we cannot gain any knowledge about the world (there is one sort of knowledge, a-priori analytical knowledge, i.e. pure logic and pure logical deduction, that can be ascertained without any input from intuitions). This is central to Kant's destructive side; because it is through this insight that we can formulate actual, certain, limits to

[34] Kant, Critique of Pure Reason, B75/p.193-194

our (non-trivial) knowledge. Whilst trivial analytic knowledge can be established by thought alone, synthetic knowledge, the kind that actually describes the world, requires intuition. As we discussed, the finitude of sensibility puts it into contact with the world, while thought can be used freely, without any resistance from the way the world actually is. This makes thought on its own *unbounded*, but this also means that on its own, there is no *friction* with the world that thought can leverage into knowledge of the world.

This content from the senses can take two forms; either we use the forms of intuition themselves as the content - in these cases we are presented with mathematical knowledge, where the form of space can furnish us with the intuition of a triangle - or more common, our intuition furnishes us with some particular intuition of something or other. This is the content about which we can reason in the hopes of gaining knowledge from reasoning.

This shows us immediately where thinking can go wrong. It is exactly when we reason about that of which we (can) have no content given by the senses (either outer, inner, or its forms). There is an understandable tendency to do so, Kant argues. For the forms of thought, categories and concepts, more of which later, are tempting to be applied without content. It is exactly from this practice, Kant claims, that dogmatic metaphysics stems, and moreover, from whence its attractiveness stems. Kant is destructive; he argues against the possibility of certain kinds of knowledge despite our natural tendency for considering the matters.

In the *Dialectic*, Kant discusses claims to knowledge, of which we can be *certain* that we cannot know them, nor their converses. For, as Kant puts it:

> *The categories consequently have no other use for the cognition of things except insofar as these are taken as objects of possible experience.*[35]

And asserting the importance of this claim:

---

[35] Kant, Critique of Pure Reason, B147-148/p.255

> *The above proposition is of the greatest importance, for it determines the boundaries of the use of the pure concepts of the understanding in regard to objects […]*[36]

Those things that we cannot, in principle, gain intuitions of, are things we can also be certain that we cannot get any knowledge about. These topics are further worked out in the *Dialectic*. Kant for example approaches God as one of those objects that is 'empty' as is cannot be given intuited content. However, we should heed Kant when he tells us no knowledge is possible about these things; although Kant makes an effort to show the faulty reasoning in a number of proofs for the existence of God[37] the conclusion cannot be that God does not exist. Nor can the conclusion be that we know God does exist. The only conclusion we can arrive at is that we do not know, God is not an appropriate object for any non-logical knowledge of, since it is the product of the 'blind thoughts' that have not been informed by the intuition. It is in this way that he attempts to resolves age long metaphysical disputes. By showing that both sides of a metaphysical debate are incapable of proving their point, Kant argued that the entire discussion should be sidelined. Whether space is infinite or not, is not something one can argue for correctly; both sides of the argument were wrong to expect to gain knowledge about this topic, for the concepts contained no intuited content and cannot be given it either.[38] Kant is destructive in the sense that metaphysical disputes are characterized in the above fashion to be *truly misguided*.

---

[36] Kant, Critique of Pure Reason, B148/p.255

[37] Kant, Critique of Pure Reason, B618-658/p.563-583

[38] It is interesting to remark that Kant's methodology of discerning what can and cannot be intuited/observed in principle, has found its way into the (modern) philosophy of physics. Especially in investigations in the foundations of physics, regarding for example whether space can be continuously empty, we find 'unobservability in principle' arguments. An example of these is the following. Arkani-Hamed argues in *The Future of Fundamental Physics*, *"that even the vacuum has structure and dynamics"* (p.54). This is argued on the basis that quantum field theory predicts that the energy required to 'probe' or observe a smaller and smaller part of space, will eventually produce particles in said region of space. That is to say, quantum theory predicts that empty space is in principle unobservable, for the observation will itself at some point generate matter and as such, fill the space. By investigating what can in principle be observed (in the physicists' case, according to contemporary scientific theories) we can discover interesting facts *about the world*. In the last bit of the sentence we find the big difference between the physicists' approach and Kant's. The preceding argument was meant to show the untenability of continuously empty space according to the physicist, for, that which is unobservable, does not have a claim to existence. In a lot of the literature, we find that physicists and philosophers of physics, tend to interpret the boundaries of our knowledge and senses *metaphysically.* In some sense, this approach is more idealist than Kant's ever was, calling to remembrance Berkley's *esse est percipi*. This all is in stark contrast with Kant's way of dealing with what is in principle unobservable in the dialectics, where we should suspend judgement on the question whether space is continuous or not, because we must conclude that neither propositions can be an object of knowledge proper. To Kant, the limits of our knowledge remain

Section 1.3: The transcendental analytic: The deduction and the self

We will now direct our attention to more constructive elements of the first *Critique*, most importantly the chapter: *On the Deduction of the Pure Concepts of the Understanding* from the second edition of the *Critique*.[39] This chapter is preceded by *On the Clue to the Discovery of all Pure Concepts of the Understanding*[40] which are preliminary remarks about what kinds of pure concepts of the understanding *we might expect* to be able to deduct. The actual deduction is followed by the part *The Analytic of Principles*[41] which is about a transcendental investigation into judgement, specifically, into how we can judge certain intuited perceptions to be subsumed under categories (i.e. under general concepts that belong to the realm of thought). As Kant states it:

> *Now how is the subsumption of the latter under the former, thus the application of* ***the category*** *to* ***appearances*** *possible, since no one would say that the category, e.g., causality, could also be intuited through the senses and is contained in the appearance? This question, so natural and important, is really the cause which makes a transcendental doctrine of the power of* ***judgment*** *necessary, in order, namely, to show the possibility of applying pure concepts of the understanding to appearances in general.*[42]

These three parts make up the *Transcendental Analytic*, that is to say, the transcendental investigation into what makes *good* thoughts possible and how these relate to perception.

The Deduction of the Pure Concepts of Understanding is a notoriously difficult part of Kant's *Critique*. It's supposed function is to show how it is necessary that 1) we synthesize the content given by our senses into larger units of knowledge, starting with synthesizing it to be mine and 2) how from this

---

epistemological. (The main gist of this footnote is lifted from a paper for the course "Philosophy of Space and Time" that I wrote; *Occam's Physics*)

[39] We will exclusively look at the B-edition version of the deduction. The B-deduction is located at: Critique of Pure Reason, B129-B168/p.245-266

[40] Kant, Critique of Pure Reason, B91-B117/p.204-218

[41] Kant, Critique of Pure Reason, B169-315/p.267-365

[42] Kant, Critique of Pure Reason, B176-177/p.271-272

follows that we should expect this synthetization to coincide with the possibility of bringing the sensory input under categories.

The categories are the concepts that Kant argues are the bits and pieces we bring to the table when we attempt to attain knowledge. These categories are supposed to overcome skeptical problems; we want, for example, *causality* to be the sort of thing of which we have knowledge. Kant's way of making this possible is that it is natural to thought to *recognize* causality. Or to put it differently, what it means for us to *think about empirical content at all*, means that we *bring what is given to us under concepts.* The idea then being that the twelve categories are the primal ways to *conceptualize* sensory input. This should then show that, given that to recognize these elements in the world *simply is what it means to think about sensory input* and knowledge can only be a product of thought, we can in principle be justified in the knowledge that is justified by the existence of the categories.

What we are however very much still left with is 1) why we should expect thinking (any thinking) to simply mean bringing under concepts and 2) how we can then come to know the specific categories that are supposed to be primary in this regard.[43]

Although Kant's thoughts on problem 2) are very interesting, we will only very briefly sketch the answer here. If we know that we synthesize and bring under concepts, we can then proceed with asking: "what are we conceptualizing?". The answer is of course that we are conceptualizing the content given by the senses, i.e. intuited content. However, before, we have seen that Kant argues that that which is given by the senses *also* has a distinction between its content and its form. The forms of intuition being space and time. Recall the quote in which Kant claims that the deduction of space and time was not necessary for the use of the senses themselves. It was required however for the analysis of thought. The reason for this was that from it being given that all our empirical thinking works on content that is necessarily of some specific form, we can also know that thinking will have to conceptualize in a way that is in accordance with these forms. That is to say, the categories are the product of, and interplay between, pure thinking and the *forms of* intuition. For Kant, non-analytical knowledge is when under the categories (already interpreted into concrete categories like cause and effect) we synthesize the information we get from our senses (which are of course already been brought under the forms of intuition). These categories

[43] They are further worked out in *The Analytic of Principles*, where Kant attempts to give content to each of the categories individually. In the *Deduction* can does give us the general answer to how this is possible *at all*.

get their concrete interpretation from the fact that we know that space and time are the forms of intuition, which is the only 'stuff' thinking can take as its content.[44]

But, how do we know *that* we must synthesize, that we must *bring under concepts*? For this Kant invokes what he calls *the unity of transcendental apperception*:

> *The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing to me.*[45]

The proposition *I think* expresses that we have attained *a unity of thinking*. When we are capable of thinking the *I think*, we imply that we have brought under concepts that what we are thinking about. When I state '*I think x'* or just *'x'* while able to think the *I think* with it, we show that we have a unified account of x *for us*. First and foremost, this implies that the logical categories are applied to it; to state *'I think x'* is also to state that x "stands under the logical functions of judgement"[46].

A representation *x* that can be accompanied by the *I think* is a representation that we can in no way disinherit. To say that *I think x*, is to state my complete ownership of, and responsibility for, the representation x. But at this point, since we now accept it to be a thought thought by me, *x* must stand under the most general forms of thought, that is, logic. It is now *something for me, and the only way it can be something for me is if it at least allows for the form of* thought. We will, from the moment that we accept something to be a thought of ours, demand of the thought, that it accepts this form and thus is consistent. This means that we have attained the aforementioned unity of our thinking, the *un-conceptual* representation *x* has been synthesized under the most general form of thought and has been made

---

[44] A nice way in which Kant prefigures his critical program is in his inaugural *Dissertation on the form and principles of the sensible and the intelligible world*. Here Kant discusses the difference between composition and synthesis, the first being something that is *merely intellectual* and the second being *a proper source of knowledge*. Pure thinking can composite any number of things by use of logical *signs*. This *and* that, A v B, etc. However, synthesis occurs when there is not only a logical unity, but a deeper unity that produces *a new object from given parts*. What our thinking under categories accomplishes is a true *synthesis*, it provides objects that are not to be understood merely as an aggregation of their parts. The analysis of these synthesized objects is what allows for the attainment of knowledge. (Kant, On the form and principles, 2:387-391/p.5-8)

[45] Kant, Critique of Pure Reason, B131-132/p.46

[46] Longuenesse, I, Me, Mine, p.80

conceptual. Now Kant leverages this insight together with the claim that, given that we know all our knowledge to be mediated by the forms of sensibility (knowledge must incorporate some empirical content) we are presented with a *transcendental table of functions of empirical judgement*, i.e. the categories. Not only do we bring every thought under the form of thought in its most general form, we also (as finite beings) bring every intuition under the forms of sensibility; and this means that every conceptualization of more than *mere thought* is also already caught under both the forms of sensibility and thought.

As we introduced in 1.1, the transcendental self is the most lean conception of the self that we can formulate. The stipulative minimum for being a self is the ability to *think the I think*. This means that for there to be a self, there needs to be both *the capacity for thought*, and certain *proper content of thought*. The transcendental character of ourselves is then firstly *that we can think*, and because of that, implicitly, *that we can apply the form of thought*. This *thinking thing*, then is also capable of understanding the laws of logic, since these are the form of thought as such. Also, this thinking thing needs to be *aware of its own activity*. Otherwise it cannot think that it is *itself the thing that thinks and as such cannot think the I think*. Longuenesse calls this a *pure action-awareness*[47] (for more see footnote 131). This self-awareness and activity-awareness cannot be mediated by the forms of sensibility either (the transcendental self needs to be aware of thinking itself, as it truly is, i.e. active, not some mediated instantiation of it, i.e. passive), which means the transcendental self lies outside of the realm of phenomena. Finally, given that we are not merely such a noumenal thinking thing, but also have an empirical self (a body and a character), we also all share in the forms of sensibility and the categories. These are not part of the most lean conception of *any* self, they are however part of the most lean conception of *any self with an empirical existence*. So, for us humans, the categories and the forms of sensibility are indeed part of what it means for us to be what we are. There can be no variation among us regarding these properties. For this reason we will also count these properties as being part of the transcendental self (they cannot be changed between other empirical creatures, so it is awkward to state this to be part of our empirical existence). However, they are different from the other properties we have discussed, because pure noumenal creatures can do without these properties; so they are intimately tied up with our empirical existence. But they are disconnected from my *specific, empirical existence*.

[47] Longuenesse, I, Me, Mine, p.86

With this we have introduced the transcendental self as it functions in Kant's *Critique of Pure Reason*. We will move to explicate some issues surrounding the bodily and mental self in the next section. After that, we will broaden the scope and discuss the importance of the properties of the empirical and transcendental side of the self in Kantian ethics and the importance of our specifically split nature.

Section 1.4: Mental self? Bodily self?

What the bodily side of the empirical self consists of, will probably not be hard to form an idea about. The body a person has is on the one hand a part of himself (it is *his* body) and on the other hand clearly given empirically. This is not strange, the most intuitive way to understand 'empirical' is as something that is given to the senses, specifically senses like our eyes, ears, touch etc. And clearly our body is given to the senses, not only our own senses, but also those of others. It is this what Kant points to when he calls this perception 'outer', and precisely because this outer perception is *spatial* we can also understand that others perceive the same things as I myself perceive. The body is *spatial*, and being spatial it can be expected to be given not only to myself, but also to others. In other words, the bodily self, a part of the empirical self, is my identification of a part of that which is *given* in space as being *me*. This immediately shows us that it is not strange that we might identify different things as being ourselves in the world and that we might be *wrong* about this. One can come to the conclusion that the mirror showed you a wrong image and that you incorrectly self-identified; this is only possible because the bodily self is spatial, which is the pre-requisite for seeing anything as being *not-me*. We must identify ourselves with some object in the in the spatial-physical world (this is a transcendental necessity), but what object we decide to identify ourselves with, and what its characteristics are, is an empirical question.

That the bodily self is empirical and not transcendental is relatively easy to grasp. Much more troubling in this regard is the *mental self* as distinct from the transcendental self. Kant comments on this problem in the following way:

> *Here is now the place to make intelligible the paradox that must have struck everyone in the exposition of the form of inner sense (§ 6): namely how this presents even ourselves to*

> *consciousness only as we appear to ourselves, not as we are in ourselves, since we intuit ourselves only as we are internally affected, which seems to be contradictory, since we would have to relate to ourselves passively; for this reason it is customary in the systems of psychology to treat inner sense as the same as the faculty of apperception (which we carefully distinguish). That which determines the inner sense is the understanding and its original faculty of combining the manifold of intuition, i.e., of bringing it under an apperception (as that on which its very possibility rests).* ***Now since in us humans the understanding is not itself a faculty of intuitions, and even if these were given in sensibility cannot take them up into itself, in order as it were to combine the manifold of its own intuition, thus its synthesis, considered in itself alone, is nothing other than the unity of the action of which it is conscious as such even without sensibility*** *[…]*[48]

On the one hand, we understand ourselves as being active, our thinking is spontaneous, yet on the other hand when we understand our inner selves as being *given* to our inner sensibility, we understand ourselves as passive objects of the senses. For our body, this did not feel so strange, but in this case, is the person we are, not exactly the *I* that presents itself to our inner sense?

The answer is no, the two are distinct. We have a direct action-awareness, as we discussed in the previous section, which means that we are directly aware of what we are consciously doing or thinking. The mental self is supposed to be distinct form this self-awareness. The transcendental self is known through the apperception, where we are directly aware of our own activity; here we are shown to ourselves in our capacity of a subject. In contrast, the mental self is given through inner sense, like our body is given through outer sense. Here, we relate to ourselves in the way of a subject to an object. Our anger is given to inner sensibility, we *are affected by anger* and *come to be aware that in fact we are angry*. And, like with outer sensibility, inner sensibility is incomplete, fallible and caught within the form of time. I might be angry without noticing myself to be so, I might be sleepy without noticing myself to be so, I might have a courageous disposition without noticing to have this and I might want to take a trip abroad and fail to notice. Inner impressions, emotional affections, dispositions and desires are all part of my mental self and can all escape being properly intuited. In contrast, being aware of what you are doing in an unmediated fashion is simply to state that *given that a thought is a thought that can be made yours because it allows for itself to be conjoined with the I think, it must be present to you directly.* When you

[48] Kant, Critique of Pure Reason, B152-153/p.257

think the *I think*, you can intellectually *understand* that given that you are doing it, it must be so that you are aware of it, else you would not be doing it. The mental self by contrast is exactly the passive dimension of our mental lives; not the decisions we take, nor the thoughts we think, but the emotions we experience, the impressions we perceive, the (habitual) dispositions we have come to own and the things we feel we want to achieve.[49]

Another way of seeing the difference is by seeing that the transcendental self cannot differ between different persons, it is always identical to itself, because, being as *narrow* as possible, when anything is taken from it, or added to it, it ceases to be a transcendental self. In contrast, the empirical self is *given* to our own senses, and since we are not autonomous in what is given to us, the way the empirical self is can at all times change. All properties of a person may at some point have been the case and at another may no longer be the case, there is nothing in my *simply being a person* that pins these properties down; and there is no reason to assume any overlap between the empirical characteristics of one person and another (some people are angry, some courageous and some others are not). But, to be the sort of creatures that we are, not merely thinking things that have nothing to think about (holy creatures)[50] nor empirical creatures without the capacity for thought (animals), we need both the unity and synthetization provided by the transcendental self to allow for thought and the diverse empirical given that enters us through our empirical existence.

Section 1.5: The empirical and transcendental self in Kantian ethics

The dual structure of the self, half empirical half transcendental, plays a large role in the Kantian conception of ethics. There are two ways for Kant that our split nature is essential to the possibility of any morality at all. One is because the unmixed versions of the selves, pure empirical and pure transcendental

[49] Kant repeats this difference again in the *Anthropology from a Pragmatic Point of View*, where he wants to warn us against mistakenly taking apperception of the self for the self caught under the form of time: *"The cause of these errors is that the terms inner sense and apperception are normally taken by psychologists to be synonymous, despite the fact that the first alone should indicate a psychological (applied) consciousness, and the second merely a logical (pure) consciousness."* (Kant, Anthropology, 7:142/p.33)

[50] Kant, Groundwork for the Metaphysics of Morals, 6:456/p.72

existences, are not capable of undertaking action, the relevant moral unit in Kantian moral philosophy. Secondly, the idea of giving the law to oneself, or as Kant calls this, being autonomous, is by being a split creature, one side lawgiver and one side law-abider. Before we show this, let us briefly introduce some extra qualities of the transcendental and empirical self, which manifest themselves in the domain of action; freedom and determinism.

Section 1.5.1: Freedom and determinism

We already briefly mentioned that thought was characterized by *spontaneity*, while intuitions were characterized by their given nature. There was there an unspoken assumption that we are *free to think this or that*. In the practical domain, this unspoken assumption becomes explicitly stated. Kant is a libertarian, that is to say, an advocate of the *freedom of the will* in the practical domain of reason.[51] But this freedom cannot be localized in the empirical world, the phenomena are caught under the category of causality. This means that freedom of the will, and the will[52] itself are *parts of our transcendental selves*. At the same time, our selves that are given to intuition, the mental and bodily self, must be understood as fully determined by the causal chain. There is no freedom in the phenomenal world, and our empirical selves are thus understandable as causally determined creatures:

---

[51] Freedom of the will is often thrown together with the postulates of practical reason: *the immortality of the soul* and *the existence of God*. These postulates are often understood to be deflationary in nature; Kant himself tells us that they are merely "*subjectively necessary*" (Kant, Critique of Practical Reason, 5:4/p.4). Freedom however resists this initial deflationary attitude, at least on Kant's own view. Freedom *must be asserted*, for else we can make no sense of the moral formula that *ought implies can*. Textual evidence for this is abound, but for one we point to the different locations where they are discussed. The postulates are discussed in the *Critique of Practical Reason* in its *dialectic*, the same place we found Kant's destructive work in the first *Critique*. Freedom meanwhile is discussed in the *Analytic of the Critique of Practical Reason*, where we found in the first *Critique* Kant's constructive work. Moreover, the concepts of God and the immortal soul gain their force *"[…] by means of the concept of freedom"* (Kant, Critique of Practical Reason, 5:4/p.4) However, as we discussed before, in the *Critique of Pure Reason*, specifically the *Canon of Reason, The Antinomy of Pure Reason; The Third conflict of the Transcendental Ideas* (B472-479/p.484-489), and the paralogisms, we find Kant being far more deflationary in regard to the self and freedom.

[52] In section 3.5 we will go into much more detail regarding how this freedom is supposed to work in the Kantian framework. For now we will use a very general understanding of what a freedom of the will consists in. More important is to notice now that it is only to be placed in the noumenal realm and thus is part of our transcendental self.

> *Hence it is an unremitting problem of speculative philosophy to show at least that its deception of a contradiction rests on the fact that we think of the human being in another sense and in other relations when we call him free than when we take him, as a piece of nature, to be subject to its law, and that both not only can very well stand side by side, but also that they have to be thought as necessarily united in the same subject […]*[53]

We need to understand ourselves as exactly those creatures which in action and thought are completely free and yet in our empirical guise are fully determined by the natural order. We will delve more into the importance of this claim in 3.5, when we look at how Kant attempts to solve the problem of evil. However, the gist of the argument resides in the necessity of the possibility of doing good *and* evil. If we cannot demand the impossible, yet the natural order decides our next action and this action is in contrast with what is demanded, the responsibility of the person for her actions disappears. Kant wants to make room for the possibility of always doing the action that is correct in certain circumstances; freedom of the will pins this possibility down, for every action could always also have not been executed, and every physically possible action could be chosen. In chapter three, we will see how Korsgaard and O'Neill, neo-Kantians, attempt to argue against the metaphysical machinery Kant employs here, while retaining Kantian moral systems. Here we will expand further on Kant's own elaborate views as well.

Section 1.5.2: Transcendental and empirical action

For Kant, the relevant unit of moral evaluation is the *action*. An action is *an act* a *done for some end* e.[54] This action, and the *maxim* (the description, or representation of the action) both have an *ends-means* structure. We want to obtain a specific end, and for this we consider a certain means, or we execute a certain act. The action and its descriptive maxim, needs to be in accordance with the *Categorical*

[53] Kant, Groundwork for the Metaphysics of Morals, 4:456/p.72
[54] Korsgaard, Self-Constitution, p.11-12

Imperative, a demand put on the *form* of the end-means structure. We will discuss in 2.2 and 2.3 the specifics of how and why this is the only possible object of moral consideration according to Kant. But for now, let us see that, given this form, it is unclear that we can understand a purely empirical existence to be capable of action, and that is unclear that we can understand a purely merely transcendental self to be capable of action.

The purely empirical creature would lack the capacity for thinking the *I think* and lack action-awareness. This would mean that although the pure empirical creature might have ends and undertake acts, these cannot by that creature be conjoined with the *I think*. There is no possibility to think *I will act this way* or to think *this is my end that I would like to attain*. And this does occur to us sometimes, I might strike someone reflexively, and the psychologist might tell me that *the cause for my acting this or that way was subconsciously my fear for my father*. But in these cases, although there might be unwilled acts and subconscious ends, it is unclear I can be said to have *done unwilled act* a *for subconscious end* e. I do not accept the end as my own end (I have not conjoined the end with the *I think*) so I cannot have done anything *to attain it for myself*. For, if I do not know myself to have end *e*, there can be no sense in which I can *reason* that I should attain that end by some act *a*. The coherent unity that is an action (an act done for the sake of some end) requires more than *just the end and the act*, it requires that the two are related to each other in the actor's thought. We must understand the end to be *worthwhile* by understanding the end to be *my own end*, and we must understand the act to be worthwhile due to its being conducive to *some end that I accept as mine*. But, to understand things to be mine is exactly to (at least) be capable of conjoining the representation of that thing with the *I think*. A pure empirical creature would be incapable of making its ends its own and thus could never *do something so as to attain that end*.

However, why would we conclude that the purely transcendental does not act at all? This is not because it cannot have a reasoned account of what to do, it can *think the I think*. In the *Groundwork to the Metaphysics of Morals* Kant claims about these sorts of (holy) creatures that:

> *A perfectly good will would thus stand just as much under objective laws (of the good), but it would not be possible to represent it as necessitated by them to lawful actions, because of itself, in accordance with its subjective constitution, it can be determined only through the representation of the good. Hence for the divine will, and in general for a holy will, no imperatives*

> *are valid; the ought is out of place here, because the volition is of itself already necessarily in harmony with the law.*[55]

The transcendental self, or the person considered apart from all its empirical properties, will have nothing to stand in between itself and acting in accordance with the Categorical Imperative. For this reason *the ought is out of place*, because the only way for a holy will to act is in accordance with moral necessity.

But this leads us to the following question: Can a purely transcendental creature actually come to action? The problem the purely transcendental creature would run up against would be the formulation of ends. For, from where would these ends present themselves to this creature? End are not spontaneous, but given and are connected to our empirical existence. Because, when I want things, *my desire for x* is not something I can decide about, *whether or not* this desire is given to me. It is only after the initial givenness of my desire that I find myself in the position to generate ways of acquiring this end, or even, deciding that *I will not act on this end. So, we can decide not to act on something that has to us the status of an end, but we cannot decide for that thing not to be an end for us at all.*

An empirical creature might of course, when he understands himself as an empirical creature, undertake actions so that, in time, it is no longer a desire that presents itself to me. I might *want* a cigarette, I might then decide not to smoke a cigarette, I might undertake action to In the future not have this craving for the cigarette (nicotine patches, finding a route to work that does not pass the tobacco store, etc.) but what I cannot do is *decide that I actually have no interest in it at all*.

The only end available to a transcendental creature is an end that is demanded by being a thinking being alone. Kant asserts that this is the *end to be good*, or as he puts it, *a love for the principle of justice.* However, the problem with this end is that *justness* underdetermines action. We must combine our contingent ends together with the end to be good, because this can pick out one specific way of acting that is prevalent to the current situation. But if all I cared about was being moral, it is unclear that *one specific way of acting*, will present itself, nor is it clear that any way of acting presents itself. Without maxims presented through the formulation of ends and means, there is no chance for the transcendental self to *choose to undertake the morally acceptable ones of these maxims*.

[55] Kant, Groundwork for the Metaphysics of Morals, 4:414/p.30

Where does this leave the purely transcendental self? It leaves it completely *in-animate*. Robbed of the given character of ends, the only action that can be undertaken on its own is the action to be just. Clearly, if it were to act, the transcendental self would do so in accordance with the Categorical Imperative, but it does not. It does not undertake actions.

Section 1.5.3: Giving the law to oneself

We are free to act good or bad and we are the ones who are the ones who obligate us to do the right thing. This is why Kant insists on the importance of our *autonomy*; the moral law should be enforced by ourselves on ourselves. No external lawgiver can provide us with the moral law, we must present ourselves the moral law and in presenting it, decide to follow the law that has been justly put to us by ourselves. As we saw before, only humans appear to be the creatures capable of coming under moral scruples, because they are the only creatures that can undertake actions. For this reason, it already makes no sense for these other creatures to give the law to themselves, for they would have nothing to apply it to. However, Kant claims that the problem is even broader, the whole idea of *being autonomous* and *giving the law to oneself* is a concept that is also only accessible to our specifically split person, he tells us in the *Metaphysics of Morals*:

> *1 The Concept of a Duty to Oneself Contains (at First Glance) a Contradiction*
>
> *If the I that imposes obligation is taken in the same sense as the I that is put under obligation, a duty to oneself is a self-contradictory concept. For the concept of duty contains the concept of being passively constrained (I am bound). But if the duty is a duty to myself, I think of myself as binding and so as actively constraining (I, the same subject, am imposing obligation). And the proposition that asserts a duty to myself (I ought to bind myself) would involve being bound to bind myself (a passive obligation that was still, in the same sense of the relation, also an active obligation), and hence a contradiction. - One can also bring this contradiction to light by pointing out that the one imposing obligation (auctor obligationis) could always release the one put under obligation (subiectum obligationis) from the obligation (terminus obligationis), so that (if both*

are one and the same subject) he would not be bound at all to a duty he lays upon himself. This involves a contradiction.[56]

Giving the law to yourself is at first glance a troubling proposition. We need to make sense of two dimensions of lawgiving that appear contrary. There is the active dimension where we bind and there is the passive dimension where we are bound. These natures appear so contrary that we cannot make sense of one and the same person, in one and the same moment, being both. Nor can it make sense that the law would not bind us if we would merely release ourselves from its binding force (amounting to the absurd situation where we need to *decide whether we value morality highly, and if we don't it would no longer be a binding force on us*). To make matters worse, without this idea of giving the law to ourselves, Kant continues:

2 Nevertheless, a Human Being has Duties to Himself

For suppose there were no such duties: then there would be no duties whatsoever, and so no external duties either. - For I can recognize that I am under obligation to others only insofar as I at the same time put myself under obligation, since the law of virtue of which I regard myself as being under obligation proceeds in every case from my own practical reason; and in being constrained by my own reason, I am also the one constraining myself.[57]

The whole idea of morality dissipates, because each and every duty must, at some point, come to be recognized by myself as being relevant to me. Without this, the law that is put upon me is fully heteronomous, with no way for me to ascertain why it is that this law is actually lawful for me. But if we are incapable of understanding why a law is lawful, we can no longer be reasonably expected to adhere to that law. Is the entire idea of giving the law to oneself doomed? Luckily not, Kant goes on to solve this problem by recourse to our dual nature:

---

[56] Kant, Metaphysics of Morals, 6:417/p.185
[57] Kant, Metaphysics of Morals, 6:417/p.185

3 Solution of this apparent antinomy

When a human being is conscious of a duty to himself, he views himself, as the subject of duty, under two attributes: first as a sensible being, that is, as a human being (a member of one of the animal species), and second as an intelligible being (not merely as a being that has reason, since reason as a theoretical faculty could well be an attribute of a living corporeal being). The senses cannot attain this latter aspect of a human being; it can be cognized only in morally practical relations, where the incomprehensible property of freedom is revealed by the influence of reason on the inner lawgiving will.

Now the human being as a natural being that has reason can be determined by his reason as a cause, to actions in the sensible world, and so far the concept of an obligation does not come into considerations. But the same human being thought in terms of his personality, that is, as a being endowed with inner freedom, is regarded as a being capable of obligation, and, indeed obligation to himself (to the humanity in his own person). So the human being (taken in these two different senses) can acknowledge a duty to himself without falling into contradiction (because the concept of a human being is not thought in one and the same sense).[58]

The solution is to understand ourselves as two distinct selves that are unified in one and the same person. Our being an empirical self that is gifted with reason, i.e. a transcendental existence, is what allows for giving the law to ourselves, for lawgiver and law-follower are not one and the same self. One the one hand, in the activity of binding ourselves, our empirical character takes on the passive nature of the one that is bound. Meanwhile, in binding ourselves, we understand our transcendental character to take on the active nature for the activity of binding. This part is, as Kant says, inaccessible to the senses, and is where our *incomprehensible property of freedom* is located and is discovered by the fact that reason influences our choice. In section 3.5 we will delve deeper into the *exact* way that we provide the law to ourselves and how freedom relates to this, by discussing Kant's dual structure of a *Wille* and a *Willkür*. The *Wille* is common to all of those we can *think* and *reflect*, i.e. is part of our transcendental self, and gives the law. This law is given to the *Willkür*, our power of choice, which must bring the demands of reason into *action in the phenomenal world*, i.e. we must bring our empirical self and its acts under the command posed by the transcendental self. So, it is the specific dual-structure of our selves allows us to

[58] Kant, Metaphysics of Morals, 6:417/p.186

understand the awkward demand that we *give the law to ourselves*, "because the concept of a human being is not thought in one and the same sense".[59]

Section 1.6: Conclusion regarding the relation between empirical and transcendental self

Let us recap what we have gotten to know about the relation between the transcendental and the empirical self up to this point by an analogy. In fragment 12, the pre-Socratic philosopher Heraclitus is said to have written that[60]:

> *In the same rivers ever different waters flow – and souls are exhaled from what is moist.*[61]

And in fragment 49a:

> *We step and do not step into the same rivers, we are and we are not.*[62]

The problematic thought this should conjure in our minds is that, although on the one hand I step into the same river as the one I stepped into yesterday, on the other hand none of the water in the river that I step

[59] We must add that the English translation and the original German text are ambiguous about *what has an obligation to what, and in what sense something 'obliges' another thing*. We have opted for the above interpretation (the transcendental self obliges the empirical self) because this would be in accord with Kant's views on the *Wille* and *Willkür* which we take from other sources (most importantly *Religion within the Bounds of Reason* and Allison's *Kant's theory of freedom*). If the empirical existence would oblige our transcendental existence there is no way to make sense of Kant's claim that it is our *Wille* and our *reason* that ascertains and enforces the moral law. If one is tempted to the other interpretation, we would first invite them to read 3.5

[60] It has been argued that we should exclude "- *and souls are exhaled from what is moist"* from the fragment, but this does not need to occupy us further (Kirk, Natural Change in Heraclitus, p.36).

[61] Heraclitus, Fragments, B12/p.70

[62] Heraclitus, Fragments, p.70/B49a

into the next day, is still the same water as yesterday. It is the same river, but different water, so paradoxically, I cannot step into the same river twice.

Although often claimed to be an illustration of, or argument for, Heraclitus' theory of *flux,* i.e. the theory that everything always changes[63], some commentators read the fragments differently. Often this is done because the idea is that Heraclitus either does not have a theory of flux that is so encompassing[64], or because the fragments should exemplify his theory that *underneath all oppositions lies unity.*[65] Another commentator, Barnes, argues that we should the understand the fragments as stating *both* the unity of opposites and that everything is always undergoing change.[66]

This reading seems sensible, for on the one hand we do not step (due to change) *but* on the other hand *do* step, into the same river twice. There are different waters, *but also* the same river. There is both a component of change, continuous flux, and a component of rest or *constancy* to the image here presented; there is a unity underlying this opposition. This means that the seemingly *solid* river is actually in a state of flux. It was a part of the river, the relatively constant *riverbed,* which suckered us into assuming a constant river. Yet, from the flux of the water, and the constancy of the riverbed, we precisely retrieve the river as a constant unity; the underlying opposites and the flux of the water in the end provides us with the unity that enables us to speak of the river as a subject that allows predication.[67]

So how is this analogous to Kant's view of the self? We have seen in the preceding parts that we may say there is an empirical component to the self and that there is a transcendental component to the self. The transcendental component is always self-identical, that is to say, because we took the transcendental to be as narrow as possible, it cannot change without ceasing to be what it is; it cannot be altered. The empirical self however is subjected to continuous change, from moment to moment. Every cell of my body is replaced in a matter of months; every thought and affection in my mind is in a state of

---

[63] This reading is often stated to originate from Plato (Kirk, Natural Change in Heraclitus, p.35) when he writes in the Cratylus: *"Socrates: Heraclitus says somewhere that "everything gives way and nothing stands fast," and, likening the things that are to the flowing of a river, he says that "you cannot step into the same river twice" "* (Plato, p.120/402a).
[64] Kirk, Natural Change in Heraclitus, p.41
[65] As stated however, interpreting the fragments is notoriously difficult. We follow this reading for the sake of philosophical clarification, not necessarily for historical claims. We do not wish to turn Heraclitus into a proto-neo-Kantian.
[66] Barnes, The Presocratic Philosophers, p.45
[67] Here we use subject differently than in the rest of the thesis. Subject here refers to the Aristotelian use, where the subjects are the things which can have properties. When things are stripped from their properties, we are left with the *subject* or *substance*.

development or decay. We propose that the (unchanging) river(bed) would be the transcendental, the (always changing) gushing water the empirical. Both together are needed to constitute a person.

The empirical self requires a transcendental self that provides unity to the changing empirical self. In the same way, the river(bed) provides unity, provides the *form* to the *manifold* of the gushing water; the empirical is anchored in the transcendental. We have argued that the transcendental self requires the empirical self for it to be an actual person (and not an in-animated Kantian *holy will*), for without it, it cannot come to action. And indeed we find; we cannot step into a river at all if there is no water in the riverbed to step into.

Finally, although there may be an apparent opposition between the empirical and the transcendental side of personhood, the complete person is actually a unity; the opposition is merely apparent.[68] If we follow Kirk on Heraclitus, we may believe Heraclitus to say a similar thing.[69] The entirety of the river-water opposition, where the river *defines the bounds and form* of the water and the water *animates* the river, serves not as a genuine opposition, but serves to constitute an underlying unity; in our case, the *person*, in Heraclitus', presumably, *logos*.[70] That is to say that indeed, there is sense in Heraclitus' words, "*we are and we are not"*, for we are both everlasting and ever changing, and this apparent paradox is what it means to be an animated unity; it constitutes being a person.

---

[68] Barnes opposes the idea to state that Heraclitus wanted to point out that the opposition was merely apparent (The Presocratic Philosophers, p.60-61). We agree that we do not have to claim that there is no *loose* sense in which there is no opposition; the two elements are of opposing natures. But they are apparent in the sense that they are not in strict contradiction with each other when conjoined into one unity for that would lead to contradictions, something Barnes agrees, should be avoided when interpreting the philosopher (The Presocratic Philosophers, p.59-60).
[69] Kirk, Natural Change in Heraclitus, p.35
[70] Kirk, Natural Change in Heraclitus, p.37-38

Chapter 2: Normativity, Virtues & the Self

The challenge of accounting for virtues to Kantian philosophy

Abstract: In this chapter we will look at a number of ways that virtue ethics might relate to Kantian ethics. We first argue that it might be tempting to take a tolerant approach, where the two sorts of moral philosophy autonomously exist next to each other. However, this cannot be satisfying for a Kantian; we flesh out Kant's claim that only spontaneous and self-conscious activity can be deemed *normative*, and argue that this means virtue cannot be normative *as is*, as we are not free to be virtuous. If we want a normative account of virtue, we need a Kantian account of virtue, which we propose to understand in terms of the transcendental and empirical self; *virtue's privileged moral status is understood by the distinct relation of privileged responsibility between one person's transcendental self and their (own) empirical self.* After this, we will look for textual evidence that Kant can cope with a number of the problems virtue ethicists want to answer. Most importantly, we deal with the complaint that Kant's account of moral motivation is too rigidly married to *respect for the moral law*, by arguing that respect of the moral law is not exclusive with more *feeling based* motivation. Finally, we conclude by arguing the primacy of *striving for virtue*, over *being virtuous*, and the future-oriented nature of virtue.

Section 2.1: Virtue ethics and Kantian ethics

In recent years, some moral philosophers, most explicitly *virtue ethicists*, have grown dissatisfied with the *narrow focus* of Kantian (and utilitarian) moral philosophy. The worry is that Kantian and utilitarian moral philosophers are too concerned with the moral status of *actions*, while they leave out questions regarding the moral status of *people*, or *lived lives*. Rosalind Hursthouse in her *On Virtue Ethics* formulates the core of the troubles with the Kantian and utilitarian programs as follows:

> *There are a number of different stories to be told about why an increasing dissatisfaction with deontology and utilitarianism should have resulted in the revival of virtue ethics […], but certainly part of any story seems to be that the prevailing literature ignored or sidelined a number of topics that any adequate moral philosophy should address. Two I mentioned above – motives and moral character; others were moral education, moral wisdom or discernment, friendship and family relationships, a deep concept of happiness, the role of the emotions in our moral life, and the questions of what sort of person I should be, and of how we should live.*[71]

[71] Hursthouse, On Virtue Ethics, p.2-3

This 'core' of the troubles is evidently quite broadly sketched here. Not all of the problems seems as primary; motives, moral character, what life to live and who to be seem to take a primary role in the dissatisfaction. Another place we find a clear sketch of what is so unsatisfactory about current moral philosophy can be found in Charles Taylor's *Sources of the Self*:

> *Much of contemporary moral philosophy, particularly, but not only in the English-speaking world, has given such a narrow focus to morality that some of the crucial connections I want to draw here are incomprehensible in its terms. This moral philosophy has tended to focus on what it is right to do rather than on what it is good to be, on defining the content of obligation rather than the nature of the good life; and it has no conceptual place left for a notion of the good as the object of our love and allegiance […]*[72]

Again we are presented with a wish to incorporate into moral philosophy the concepts of *being a good person* and *living a good life*. Added to this is in another form the question of moral character and motivation; Taylor wants a way to relate to the good in a way not reducible to mere *acceptance, or understanding of duty*, but also as being motivated by the good, by a *love for the good*. Something that goes deeper than mere rational acceptance, but ingrains itself into one's person, providing the person as something that *harbors the good*.

What Hursthouse and Taylor at least seem to agree on is that we not only want an answer to the question what to do, but also who we should be. And the answer to this question, at least on Hursthouse's view, is of course that we need to be virtuous. Virtues imbue our character with moral value (to have a courageous character is *better* than to have a cowardly one), give a view of the proper motivation for actions (we should enjoy, and happily undertake, good actions), provide an idea of moral education (the breeding of virtue) provide proper emotional responses (emotional reactions can be understood as the

[72] Taylor, Sources of the Self, p.3

same sort of dispositions as virtues are understood to be) and of course *tell us who we should be*. The only question that does not fit as well with virtues is that of *the good life*. We will sideline this issue.[73]

Virtue then is a powerful antidote to some of the perceived problems with utilitarianism and Kantianism. As Kantians we can of course wonder: *are we capable of incorporating an account of virtue in our theory*? Hursthouse herself celebrates and encourages these kinds of projects.[74] Hursthouse's enthusiasm can be curbed by a counter-question however: is there any reason to take a specifically Kantian approach to virtue. *Why not simply take virtue ethics as its own, distinct, autonomous philosophical activity?* On the first question the answer is decidedly 'yes'. Modern Kantians like Onora O'Neill[75] and Christine Korsgaard[76] go out of their way to incorporate accounts of virtue into their theories. But also in Kant's own texts we can find clues and in some cases, systematic philosophizing, on virtue.[77] The second question is more tricky. The free transcendental self and the causally determined empirical self as the model for the subject of action is introduced by Kant for his moral theory because he claims this is *the only possible form a moral theory could take*.[78] If there is truth to this Kantian claim, then perhaps there is no sense to an autonomous moral philosophy of virtue, for there can be no moral philosophy not in some way having taken on the Kantian guise, and in turn, this behooves us to consider all the more, virtue based extensions to the Kantian program.

But, to claim that the only possible way a moral philosophy could be, is essentially Kantian, is quite a bold claim. As Charles Taylor rightly points out, this reeks of semantical exclusion from the predicate 'moral', an exclusion that tells us quite little about the nature of virtue ethics as an autonomous activity:

---

[73] The good life is very hard to pin down and resists philosophical theorizing because whether a lived life was a proper life seems to be tied up not only with the specifics of one or other life, but also depends on the understanding the one who lived had of this life. In 4.4 we will touch upon the idea of *narrative exercise* as way to change, or imbue, one's life with the *telos* that would make the life a well lived one were it to be attained. Still, it is far from clear this is an answer on the *idea of the good life*, for it is merely an understanding on how we can come to understand ourselves as being tied up essentially with one or another program of action.

[74] Hursthouse, On Virtue Ethics, p.3-4

[75] *Justice and Virtue* explicitly delves into what O'Neill perceives to be an improper opposition that is presented to be the case between virtue ethics and Kantian justice based approaches to political philosophy.

[76] In *From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action* in her book *The Constitution of Agency,* Korsgaard provides a comparison between Aristotle and Kant and argues that the differences that immediately make themselves apparent can be resolved..

[77] Most telling his that he has a *Doctrine of Virtue,* which makes up half of his *Metaphysics of Morals,* which we will be discussing a little further on in the chapter.

[78] Kant, Groundworks for the metaphysics of morals, 4:393/p.9
This claim permeates most of Kant's moral writings, asserting that a Categorical Imperative can be the only Imperative of a moral nature.

> *I spoke in the previous paragraph about our 'moral and spiritual' intuitions. In fact, I want to consider a gamut of views a bit broader than what is normally described as 'moral'. In addition to our notions and relations on such issues as justice and the respect of other people's life, well-being, dignity, or questions about what makes our lives meaningful or fulfilling. These might be classed as moral questions on some broad definition, but some are too concerned with the self-regarding, or too much a matter of our ideals, to be classed as moral issues in most people's lexicon. They concern, rather, what makes life worth living.*[79]

So, can we not easily reconcile Kant's idea that his sketch of a moral theory can be the only possible one while still retaining other questions and areas of philosophical inquiry, but then under the label "morality in a broad sense"? But then, why should we care about this broader inquiry that, while left possible, may become devoid of intrinsic value? Taylor aptly continues the previous paragraph as follows:

> *What they have in common with moral issues, and what deserves the vague term 'spiritual', is that they all involve what I have called elsewhere 'strong evaluation', that is, they involve discriminations of right or wrong, better or worse, higher or lower, which are not rendered valid by our own desires, inclinations, or choices, but rather stand independent of these and offer standards by which they can be judged.*[80]

This description of *spirituality*, or *strong evaluation* by Taylor here, quite closely mimics a concept that in recent literature has been growing in popularity; normativity. Normativity is a property not only of the moral domain (although it is most certainly its most prominent instantiation) but is also ascribed to, for

[79] Taylor, Sources of the Self, p.4
[80] Taylor, Sources of the Self, p.4

example, epistemology[81], logic[82], aesthetics[83] and scientific methodology[84]. To be a normative domain is to be a domain in which there are norms that delineate good instances in that domain from bad instances. For example, the rules of logic provide a norm by which we can decide how good a certain judgement is and the Categorical Imperative provides a norm by which we can decide how good an action is. In her seminal work *The Sources of Normativity*, Korsgaard introduces the normativity of ethics (and other domains) as follows:

> *The answer lies in the fact that ethical standards are normative. They do not merely describe a way in which we in fact regulate our conduct. They make claims on us; they command, oblige recommend, or guide. Or at least, when we invoke them, we make claims on one another. When I say an action is right I am saying that you ought to do it; when I say that something is good, I am recommending it as worthy of your choice. The same is true of other concepts for which we seek philosophical foundations. Concepts like knowledge, beauty, and meaning, as well as virtue and justice, all have a normative dimension, for they tell us what to think, what to like, what to say, and what to be. And it is the force of these normative claims – the right of these concepts to give laws to us – that we want to understand.*[85]

Common parlance points our attention to many different forms of normativity. A plethora of norms, for a plethora of domains of human activity. Of course, whether each and every one of those domains *actually* contains non-illusory norms (that contain *force* as Korsgaard calls it) is a question that needs further inquiry. But in any case, there does not seem to be a *prima facie* unity to all these norms besides their being normative.

Kant argues that his moral system is the only form a moral system could take. But, should we then not just provide for a more broad understanding of morality, that does incorporate virtue ethics as its

[81] What should we believe? What judgements are proper? We find this concern go as far back as Plato's *Theaetetus* where the third part (justification) of the common definition of knowledge *true justified belief* is added because this is what makes holding specific beliefs commendable or not.
[82] What are proper derivations? How should we reason? Should we conclude from A and A -> B that B?
[83] What are beautiful works of art and how should we appreciate them?
[84] What are good ways to structure our search for empirical knowledge? What constitutes good science, and what bad science?
[85] Korsgaard, Sources of Normativity, p.9

own, separate, domain? Virtue ethics would be a domain with its own norms, like strict morality contains its own norms. We will argue however that when Kant's tells us that his moral theory is the only form a moral theory could take he is *not* talking about strict morality. *Kant's terminology undermines the understanding of his claim as being essentially an analysis of normativity*. If this is so, we must with renewed vigor take up the issue of virtue within a Kantian scheme.

Section 2.2: Norms and reason

In the previous section we have seen a few properties of normativity pass. For a domain to be normative, it must: i) contain norms which decide some concrete instances in the domain to be *commendable* and other to be *reprehensible* ii) be possible to explicate the *force* of these norms iii) place these norms on people irrespective of people's contingent desires, inclinations, and knowledge (Taylor calls this a requirement of *independence*) and iv) expect the people whom these norms are placed upon to adhere to them.

We will see that it are the third and fourth properties that Kant cashes out, to argue that his own account of morality is the only possible one. But, normativity is not a concept merely applicable to the moral domain; to help understand Kant's argument on normativity in the moral domain, we will make a parallel between the practical domain and the theoretical domain. So, before we move to his analysis of the only possible demand that we could put on our actions as expounded in the *Critique of Practical Reason*, let us return to Kant's claim in the *Deduction* that:

> *The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible* ***or else at least would be nothing to me***.[86]

[86] Kant, Critique of Pure Reason, B131-132/ p.246

This statement is not just Kant's de facto description, or stipulative definition, of our constitution as thinking beings. It is a claim about the status of certain representations for a thinking being, some being something to me, and others, when impossible to be accompanied by the I think, being nothing to me. Note that Kant rebounds from the claim that a representation that the *I think* could not accompany, is truly impossible; it is not exactly impossible, it simply is nothing to *me*. We propose to read in this a Kantian analysis of normativity. Say some representation would occur within me that I would *not* be able to accompany with the *I think*. Furthermore, say that the representation fails to conform to some epistemological norm: a false judgement, an illusion, etc. In such a case the epistemological norm is not a norm for me; incapable as I am to *make the representation mine by thinking 'I think this representation x'* the norm fails to adhere to the fourth property we demanded from it (the norm is placed on some representation, but not on me, for I am incapable of making the representation mine). Any representation that truly cannot be accompanied by the *I think*, we take Kant to tell us here, *is devoid of normative power, as it is nothing to me*.

But why does Kant provide this privileged place for *self-conscious thought* (or the capacity for thinking *I think*) in relation to normativity? This because we cannot possibly understand our self-conscious thoughts as being distinct from us. The moment we state *I think x*, we take ourselves to be responsible for the content of the representation *x* as we have conjoined it to our spontaneous capacity of thought. We neatly fall into place with that which we understand to be up to us; we can decide, once we are aware of the representation x, what we *think* of the representation x.

And this is not merely an investigation into the nature of proper judgement or thought. This point is well taken for any *norm put to us*. If the norm is a norm for things that we cannot take control of ourselves (that we cannot make *ours)*, it cannot be a norm for us at all, and as such, can even be argued to not be properly normative in any case. This is why Kant introduces his normative demands for thought and actions in relation to critiques of *reason*. It is only insofar as we understand the adherence to norms to be up to us, that we can accept them as norms for us. But this means that norms are restricted to norms posed by reason itself, as *we are only then at all times capable of grasping and thus adhering to them*, as Korsgaard puts it:

> *What obligates me is reflection. I can obligate myself because I am conscious of myself.*[87]

Norms from the outside are problematic (for how do I know this external object without fail), and norms 'norming' things outside of us (or simply *not norming us*), are equally troublesome, as these objects normed are not up to us i.e. are not ours at all. Norms can only come from reason, and can only concern self-aware activity (which cannot be understood as not being up to us).

From this, we can quite readily reconstruct Kant's insistence on *actions* as being the locus of morality. What's more, we can also reconstruct why it must be the highly formal *Categorical Imperative* that is allowed as the supreme principle of morality. All of this returns not only to a specifically Kantian approach to *morality* but to his analysis of *normativity as such*.

Firstly, actions must be the locus of moral philosophy, because *we self-consciously act*. What's more, in Korsgaard we find a move that narrows down actions to only those normatively relevant in a similar fashion as did Kant for representations:

> *Let me introduce some terminology in order to express these ideas more clearly. Let's say that the basic form of a Kantian maxim is "I will do act-A in order to promote end-E." Call that entire formulation the description of an action. An action then involves both an act and an end, an act done for the sake of an end.*[88]

Actions must be done consciously, combining both means and ends into a maxim. An end that we do not act on clearly resist normative evaluation. But even an act done without an end resists normative evaluation. This is because an act devoid of an end is not a consciously undertaken act. Actions are the units that are understood by Korsgaard to be essentially self-conscious, like the representation that is accompanied with the *I* think is a self-conscious unit for Kant in *The Critique of Pure Reason*. We cannot shy away from our responsibility for our actions, because given what they are, *they were completely up to ourselves, as they were self-conscious*. We decide to provide a means to some given ends. Only actions

---

[87] Korsgaard, Sources of Normativity, p.136
[88] Korsgaard, Self-Constitution: Agency, Identity and Integrity, p.11

can be 'normed' in the practical domain because only actions are *spontaneous*, like thinking is in the theoretical domain.

Not only must we be able to follow the norm, we must also ourselves be capable of ascertaining the norm. A norm for me that cannot be known by me, cannot be a norm that binds me. This captures the third property of normativity we found; the norm must be irrespective of my specific inclinations and desires. Even more, it needs to be irrespective of my concrete knowledge of the world. This means that knowledge of what the norm is must be ubiquitously accessible to all of those who are to be bound by the norm. So, it cannot be provided through external means, for those might be inaccessible for some. It must be implicit in my *being the sort of creature that is to be bound by this norm*. This is why Kant insists on the normative force provided by reason alone. For, only those creatures with reason can be expected to adhere to the norm and at the same time, all creatures which are reasonable are then capable of finding the law of reason within themselves.

This leads to the following analysis in the *Critique of Practical Reason* of what it is about actions that we should provide a norm:

> *If a rational being is to think of his maxims as practical universal laws, he can think of them only as principles that contain the determining ground of the will not by their matter but only by their form.*
>
> *The matter of a practical principle is the object of the will. This is either the determining ground of the will or it is not. If it is the determining ground of the will, then the rule of the will is subject to an empirical condition (to the relation of the determining representation to the feeling of pleasure or displeasure), and so is not a practical law. Now, all that remains of a law if one separates from it everything material, that is, every object of the will (as its determining ground), is the mere form of giving universal law. Therefore, either a rational being cannot think of his subjectively practical principles, that is, his maxims, as being at the same time universal laws or he must assume that their mere form, by which they are fit for a giving of universal law, of itself and alone makes them practical laws.*[89]

[89] Kant, Critique of Practical Reason, 5:27/p.24

The Categorical Imperative (the norm for actions) checks the form of our maxims. It must be the form of our maxims, not our ends or our means separately. Because these depend on concrete knowledge of the world and having concrete individualized desires. Neither can be expected to be available to all merely by being the creatures we are. The form however, of a *given maxim* that we are now considering, this is accessible to all who are considering the maxim. This is so because the maxim itself is fully self-conscious and so is open to analysis. Only the maxim as a whole, and in effect, only the form of the maxim, is properly normative. Any being that acts at all, that self-consciously undertakes acts, is also capable of reflecting on what it is doing, by definition. But, being capable of this reflection, is exactly to say that we can abstract from our specific inclinations, desires and knowledge of the world.

Then, Kant's claim that his moral philosophy is the only possible one is not mere rhetoric nor semantic playfulness. All normativity must be bound up with that *which is up to us*, and as such, all normativity in the practical domain must be bound up with *self-conscious action*. What's more, the only form a norm on actions can take is as an analysis of the form of the maxim underlying the action; for only then can we be sure that all relevant information for knowing what the norm is, is accessible to all those who are bound by the norm. Whether or not we follow Kant in this devastating critique of all normativity that is derived from something outside of reason and applied to things outside of our control, it shows at least that it is for a Kantian prima facie not possible to simply let virtue remain its own, normative, domain, disassociated form broader Kantian moral philosophy.[90]

---

[90] In our sketch of virtue ethics we provided philosophers whose greatest pains with act-centered moral philosophy is that they *miss* a number of concepts in their theorizing that seem to be essential to any complete account of morality. In this case the question becomes; can we practice virtue ethics divorced from broader ethical theories or should we incorporate virtue ethics? We answered that for Kantians it is prima facie problematic to have one's account of virtue divorced from actions, as doing so makes a mystery of the normative character of virtues. But some might have found our initial sketch of virtue ethics problematic, because other authors jump to their minds, which from the outset had a far less confined reason for advertising virtue ethics. One of the prime examples is G.E.M. Anscombe in her article *Modern Moral Philosophy*. Here virtue is not introduced as a supplement to modern moral philosophy, but as a straight up replacement; called for not by the incompleteness of modern moral philosophy but by its inherent incoherence. We will argue however, that Anscombe misses the mark with her combative attitude; to do so, we will look at the three arguments she poses against the Kantian scheme. Anscombe's first argument is as follows:

"Kant introduces the idea of "legislating for oneself," which is as absurd as if in these days, when majority votes command great respect, one were to call each reflective decision a man made a vote resulting in a majority, which as a matter of proportion is overwhelming, for it is always 1-0. The concept of legislation requires superior power in the legislator." (Anscombe, Modern Moral Philosophy, p.2)

The last remark rings untrue. It is unclear why only the one with more power can legislate, claiming this is to assume upfront that legislation can only be *de facto* binding. But even when taken as a description of legislative functions in the world it seems off. Dutch authorities legislate that I should drive on the right side of the road. But even then, I

Section 2.3: A problem of normativity for virtue and its (dis)solution in the dual-self

We started 2.2 with four properties that the norms in some domain of activity needs to adhere to, to be counted as normative. Virtues may remain untroubled (at least on our previous analysis) by the first two constraints. The norms, or the virtues, clearly describe the good *person* as distinct from the *bad* one. The force of the virtues may conceivably also be provided, in a worst case scenario, by reference to common sense understanding of the goodness of the person. However, virtues are not clearly distinct from my own specific place in the world and from my particular existence. For, whether I am *actually* virtuous is exactly a statement *about my particular mode of existence*. But, although whether I am virtuous *is* dependent on my actual desires and inclinations (one of the big problems virtue was supposed to solve was that of

---

am still capable of driving on the left side of the road (sometimes without any repercussion). Their legislative powers do not come from their mere capacity to *enforce* the legislation (which they at times lack) but from acceptance of the law by me. The talk of an 'internal vote' is also a misnomer, self-legislation is not in any form a democratic process, and my decision to not heed the legislation my reason presents me with is not an overturning of the proposed legislation, but a transgression of that legislation. That is not to say there are no troubles with legislating oneself (namely the apparent identity of legislator and legislatee) but we have seen Kant himself deal with this problem in 1.6 by recourse to the dual structure of the self. The problems introduced here by Anscombe with self-legislation however have no traction.
Secondly, Anscombe introduces the oft cited problem that Kant's account is both too formalistic and too rigoristic. (Anscombe, Modern Moral Philosophy, p.3) Maxims are, on Anscombe's view, describable in many different ways, and to give both any content (against formalism) and any flexibility (against rigorism) finding the right description is essential. But, there is no good way to do so Anscombe argues and most assuredly, Kant provides no such possibility. However, maxims come on the Kantian account *before* the action. An action is as we discussed, essentially self-conscious, as such we cannot claim that we are unaware of the actual nature of our maxim. It is true that *somebody who steals* may have had a plethora of maxims. However, for the person who undertook the action, he must *undertake act x for purpose y*, and be well aware of doing so. This means that there is no question of *redescribing* the maxim as Anscombe would have it; there is a *correct* description, the one the actor was aware of when he formulated his maxim and undertook his action. Further description can only occur in, what O'Neill (Acting on Principle, p.247) calls the *context of assessment*, but rediscription is resisted in the *context of action*.
Finally, there is a historical argument which argues that*: "The answer is in history: between Aristotle and us came Christianity, with its law conception of ethics."* (Anscombe, Modern Moral Philosophy, p.4) However, now that we no longer believe in a lawgiving God, we can make no sense of law-based ethics anymore either, despite it having rooted itself into our common parlance: *"Naturally it is not possible to have such a conception unless you believe in God as a law-giver; like Jews, Stoics and Christians."* (Anscombe, Modern Moral Philosophy, p.5). This is a very bold claim; to state that the only lawgiver that we could make sense of is an all powerful God stands to reason. The Kantian program is exactly to make sense of *reason* as lawgiver; carried within each of us in the form of our transcendental existence as reasonable beings. Kant provides the idea of self-conscious lawgiving of the self by the self; a lawgiving clearly distinct from a conception of God.
We believe that Anscombe's combative account of virtue ethics is both faulty and should not be misunderstood for the more moderate view that Hursthouse endorses. The value of virtue should not have to depend on a criticism of the other moral theories, virtue is clearly an important part of our moral lives and as such demands philosophical reflection.

having *the right* emotional responses and to act with *proper* inclinations, and to in the end *harbor the good*, all of which clearly depend on my desires and inclinations being *proper* as well) the virtue ethicists may circumvent this by pointing at the idea that *that the virtue is good to have is independent of my desires and inclinations*.

We are left however with the fourth constraint, that the norm needs to be a norm for me, and as such, needs to be a norm for self-conscious activity. For if it is to be separated from self-conscious activity, adherence to the norm is not up to us and thus not a norm on us, at all. But, having virtues, and as such adhering to the norms for being a good person, is clearly *not* up to me. Whether I rebound at the sight of blood, showing off my vices of cowardice and excitability, is a property of me that I do not get to *will away*. It is a property tied to my empirical self, the mental self, which means it is caught under the category of causality. My body and my mental life are not up to me, as they are caught in the forms of sensibility and as such, in the causal chain of the natural order. At this point, we must accept that changing parts of the natural world requires activity that is beyond the mere power of the will. Like I cannot will my body fit, I cannot will my mind courageous, this requires practice and habit formation. Virtue cannot be normative *as is*, if there is some way to make sense of the normativity of virtues, it is by a detour through action.

Virtues are properties of the empirical side of the self and as such is to some extent out of the reach of the transcendental self. Actions, justifications, derivations, these activities are part of the transcendental self, that is to say, of self-conscious activity. So the question about how we can make sense of the normative dimension of virtue can be reformulated as: *How can we make sense of the normative status of the empirical self*.

At first glance, the hard conceptual line that is drawn between the empirical self and the transcendental self seems to be of little use to us here. Does it not only make the gap between the unproblematically normative and the un-normative ever more great? But this is a problem not of the conceptualization, but as we discussed before, of the nature of virtue itself. Yet, it does point us into the direction of the solution. *Virtue should not be understood as activity, but as a settling down of activity into the self*. Nor should we expect virtue to point us to a moral object different in kind from those of actions and objects in the world. *Virtues are objects in the world*. But, the generation of virtues in the world in *our own minds and bodies* is something that *we* are more responsible for than others. This is the big way in which virtues find their own distinct moral status. Objects in the world are shared in between all of us, and to bring about a pleasant world for all of us, is a responsibility *for all of us*. However, the development of one's mental self and bodily self, lay more exclusively with the persons themselves. That is to say, the

self-conscious activity of some person (the activity of the transcendental self) has privileged responsibility for its own bodily and mental self. This makes understandable that we are more inclined to look to virtues when looking for morally relevant objects than to other objects in the natural world, as we can prima facie easily identify a responsible party for the state of affairs of the object.[91] Yet, their normative character can only be understood by someone *attempting*, or *striving* to make a proper state of affairs obtain for that part of the natural world which he understands as his own. So the normative status of the empirical self needs to be understood in terms of the person's responsibility for attempting to bring about a virtuous empirical self.[92] *The normativity of virtues derive from their roots in our actions (we must act such and so to attempt to bring about virtue) and their privileged moral status is understood by the distinct relation of privileged responsibility between one person's transcendental self and* <u>*their*</u> *empirical self.*

Section 2.4: What is Kantian virtue? Does Kantian virtue do what it should?

We have provided a basis for the strange *in between* position in moral philosophizing that virtue is to take. But, is there now in Kant a worked out account of virtues that we can work with? An account that allows answering some of the questions that we found Taylor and Hursthouse ask at the beginning of this chapter? To see this we will look at Kant's account of virtue in a number of his works (which take two distinct approaches to understanding virtue) and argue for a number of these questions that they are decently accounted for. We will start with the blunt of Kant's account of virtue, which focusses on virtue as the *predisposition to act well*. After that we will look at Kant's more broad approach to virtue in the *Metaphysics of Morals* and *The Religion within the Limits of Reason*, his last works on moral philosophy.

[91] That there is a prima-facie responsible party in the person herself does not mean of course that the person must herself be the only responsible party. We also look at parents, and can fault them for, through poor upbringing, breeding vice into their children.

[92] Of course, this is not to say that there are no difficulties in this project of molding oneself. As Heidegger nicely put it we are *geworfen Entwurf*, we both design and project ourselves, but we also find ourselves in this world with some given constitution and location that we cannot *simply* abstract ourselves from. We are a thrown design. We will discuss the problems inherent in designing yourself more extensively in chapter four.

Section 2.4.1: Kant on virtues in the Groundworks and the 2nd Critique

In the *Groundworks for the Metaphysics of Morals* and *The Critique of Practical Reason* we find Kant's view on virtue to be mostly concerned with the idea of *having the predisposition to do good action*. As Lara Denis summarizes Kant's views in *Kant's conception of virtue:*

> *[…] we can understand Kant's conception of virtue as the form in which a rational being with a non-holy will expresses her supreme commitment to morality: as a continually cultivated capacity to master her inclinations so as to fulfill all her duties, a capacity whose cultivation and exercise is motivated by respect for the moral law.*[93]

Virtue is a capacity to make good on our moral commitments, to turn the maxim (that we understood to be moral) into deed. Inclinations need to be subverted, and this capacity for subversion needs to be practiced. Even this *relatively thin* conception of virtue is capable of answering a number of the questions raised by virtue ethicists like Hursthouse, and the somewhat more general complaints by Taylor.

Firstly, one of the more prominent upshots of having a theory of virtue is that one can generate a theory of moral character. This is one of the spearheads of the virtue ethicists program.[94] But Kant's thin conception outlined above is very much capable to give some interpretation of what a moral character would consist in, it is exactly when we are capable of putting our good maxims into actions, i.e. when we have the "morally good disposition"[95].

A more involved answer we would like to give is to the question posed about having the proper moral motivation for you actions. To this Kant of course has a very well-known answer:

> *What is essential to any moral worth of actions is that the moral law determine the will immediately. If the determination of the will takes place conformably with the moral law but only*

---

[93] Denis, Kant's conception of virtue, p.513
[94] Hursthouse, Rosalind and Pettigrove, Glen, Virtue Ethics
[95] Kant, Groundworks for the Metaphysics of Morals, 4:435/p.53

> *by means of a feeling, of whatever kind, that has to be presupposed in order for the law to become a sufficient determining ground of the will, so that the action is not done for the sake of the law, then the action will contain legality indeed but not morality.*[96]

The right motivation is, in the end, the respect for the moral law that we can ascertain through the exercise of merely our own reason. For, when respect of the moral law would not be capable of motivating us, there is no way of universalizing the maxim (peoples particular motivations are private and diverse) and imbuing it with moral quality. Even the cultivation of our *thin* virtue must also be done from an understanding of its moral necessity, i.e. the cultivation of our disposition to do good, must also be motivated by the understanding of its moral necessity. So, Kant's answer to the question what is moral motivation at first seems to be that *you can have only one motivation in action, respect for the moral law.*

No virtue ethicist would be impressed by such an answer and rightly so. This is an example where there might be *an answer* to the posed question that can be formulated in the Kantian framework, but despite this, it is not answer we should be satisfied with. Luckily, we hold, that this position is not true to Kant's own views, and merely an, understandable, misinterpretation.

Firstly, let us propose how we would understand the above citation. Kant tells us that the law must determine our will. And, he also points out that, the law needs to be, of itself, sufficient to motivate our will, and that no other motivations may play a role in its being sufficient. This is, however, not the same as to claim that there can be no other motivations at all. What Kant is doing is proposing a counterfactual; if I were to strip this person of all her personal feelings and motivations, would he still be motivated to do the right thing. The answer to *this* question needs to be yes. But this does not mean that a person who does not act purely on the basis of respect for the law, but finds in himself a mixture of both respect of the law and other feelings and motivations, is then necessarily immoral.

Nor should we take it to be so that Kant here claims that for a person to be moral, the person must necessarily be motivated by the respect for the law either. We do not read Kant claiming here that *if you would take away the respect for the moral law as a motivation, you should also no longer be motivated to do the action*. No, we can be sufficiently motivated both by our empirical considerations as well as our purely moral ones. Overdetermination is often looked at as a problem to be solved, a puzzle

[96] Kant, Critique of Practical Reason, 5:71-5:72/p.60

to be cracked; which of the determinants is the *actual* one? But, for Kant, overdetermination is an essential part of his philosophizing, not a problem, but a solution. His acceptance of both noumenal freedom and phenomenal determinism are to be understood as an overdetermination of human actions. For, human action is both caught in the forms of causality[97] (and will accord with them, being determined by other mechanical forces) and is determined by the noumenally free agent's decisions. Both are sufficient, neither is necessary, and both occur.

This shows why we believe Kant would not respond that the only motivation should be respect for the law. And luckily so, for the virtue ethicists are demanding more than that with good reason. They don't necessarily want to do away with the respect for the law, but they are asking, what kinds of concrete motivations we should harbor, and say something about the moral value of *those* motivations. An example.

Two brothers are discussing philosophy, and the first, a Kantian, tells the other that: "I would never kill you, my brother, for the moral law that I respect would forbid it". The other brother is aghast, and responds: "I do not need the moral law to never be brought to kill you brother and I find it troubling that you would require the moral law to forbid you to kill me". The Kantian brother is taken aback an asks: "But, even if we were not brothers and friends, would you not through mere moral reasoning be stopped from killing me?" The second brother replied staunchly: "I do not know, I do not know how I would know, and I do not care so much, for we are brothers and friends".

We think that the virtue ethicist has something like this in mind: in this case, the second brother is right that the Kantian brother is morally troubling if he truly needs the moral law for not killing his brother, for he should have (cultivated) the right virtues, feelings and in effect motivations, that would (over)determine the issue in such a way that pure moral reasoning itself would no longer be necessary. We believe to have shown that the Kantian account of virtue outlined above at least *leaves room*, through the idea of overdetermination of the will, for the importance of having proper motivations that do not coincide with respect for the moral law. It is however in the following section that we will find an account of virtue that *states* that we should aim for other motivations.

[97] At the very least this is the case for the physical representation of human action, that is, the movements of humans in the world.

Section 2.4.2: Kant on virtues in Metaphysics of Morals and Religion within the Bounds of Reason

While Kant's thin understanding of virtue is helpful for stating its general moral import, it leaves out exactly the things that the virtue ethicist would find so important to take into account. In the *Metaphysics of Morals* Kant becomes more willing to discuss the wider application of the concept of virtue. This is based on his idea that there are three predispositions that provide motivations, the *animal* (emerging from the bodily self), the *humane* (emerging from the mental self in conjoinment with the existence of others around us) and the *personal* (emerging from the transcendental self).[98] As Denis puts it:

> *Although the predisposition to personality is obviously morally crucial, all these predispositions are good, all encourage compliance with the moral law. The drives of humanity and animality, however, are susceptible to corruption. The drives of animality can have grafted onto them vices such as gluttony and lust. The drives of humanity can have grafted onto them vices such as jealousy, rivalry, and malicious glee. Thus, a large part of the struggle for virtue is the effort to harmonize these three predispositions so that personality develops fully, and humanity and animality develop in ways supportive of morally practical reason.*[99]

If we follow these remarks, we find a more *full* account of virtue. Because, not only do we need to make sure *we have a disposition to do the good*, we now see that this is not a merely noumenal activity. Sure, personality needs to be developed fully (which would coincide with a thin conception of generating a disposition for the good) but we see that we also need to make sense of harmonizing humanity and animality with personality. This means that in becoming the sorts of people who harbor the good, we need not be only concerned with morality continuously and all the time. We see this claim already prefigured very generally in the *Groundwork*, when Kant tells us that one of the implications of the Categorical Imperative is that:

[98] Kant, Religion within the Bounds of Reason, 6:26-6:27
[99] Lara Denis, Kant's Conception of Virtue, p.515-516

> *Third, in regard to the contingent (meritorious) duty toward oneself, it is not enough that the action does not conflict with humanity in our person as end in itself; it must also* ***harmonize*** *with it. Now in humanity there are predispositions to greater perfection, which belong to ends of nature in regard to the humanity in our subject; to neglect these would at most be able to subsist with the preservation of humanity as end in itself, but not with the* ***furthering*** *of this end.*[100]

There is an imperfect duty (that is a duty that should generally be acted upon, but does not require strict, continuous, adherence) to improve ourselves. This improvement cannot be understood to be merely intellectual; we must not only improve ourselves so as not to be in conflict with our humanity, but we must attempt to improve ourselves so that *we* become harmonized with our humanity.

So let us return to the issue of proper motivation. We claim that Kant should and would agree with the lesson we draw from the discussion of the brothers. It is indeed, when considering to kill your brother-friend, not preferable to have to let the predisposition to personality, purely moral considerations, tell you not to kill. It is better instead to have harmonized your predisposition of humanity so as to implicitly know or understand that such a thing is wrong and thus not need, as Bernard Williams called it, *a thought too many*.[101] But, the Kantian does not have to accept that he needs to have a thought; the virtuousness that he has, over time, imbued into his empirical character, into his animality and his humanity, can overcome these kinds of problems without recourse to pure reason. This is acceptable because the breeding of this virtue, this harmonization of one's predispositions, is done with the demands of reason in view, just so that in the moment we can live without invoking it.

Another issue that Taylor point out is the wish for us to become creatures who *love the good*, loving laying beyond mere rational acceptance. In general, love and emotional attachment, seem to be poorly worked out themes in Kantian ethics. As we showed before, it is a misrepresentation to say that Kant wants us to repress all our emotional attachment in the favor of rational consideration. Although reason is the final touchstone, the place where all normativity derives its force from, it is not the only way that we, as creatures, can strive to become good.[102] But, how can we make sense of harboring a love of

[100] Kant, Groundworks, 4:430/p.48
[101] Williams, Persons, character and morality, p.18
[102] And of course we should not reconstruct this as Kant telling us that *it is reasonable to love this or that person*. Reason can only tell us that it is reasonable or not, to strive to defend the people we love innately. (Kant, Metaphysics of Morals, 6:448/p.213)

the good as Taylor proposes? We believe a Kantian answer would draw upon the following (rather lengthy) citation:

> *The rules for practicing virtue aim at a frame of mind that is both valiant and cheerful in fulfilling its duties. For, virtue not only has to muster all its forces to overcome the obstacles it must contend with; it also involves sacrificing many of the joys of life, the loss of which can sometimes make one's mind gloomy and sullen. But what is not done with pleasure but merely as compulsory service has no inner worth for one who attends to his duty in this way and such service is not loved by him; instead, he shirks as much as possible occasions for practicing virtue. With regard to the principle of a vigorous, spirited, and valiant practice of virtue, the cultivation of virtue, that is, moral ascetics, takes as its motto the Stoic saying: accustom yourself to put up with the misfortunes of life that may happen and to do without its superfluous pleasures. – This is a kind of regimen for keeping a human being healthy. But health is only a negative kind of well-being: it cannot itself be felt. Something must be added to it, something which, though it is only moral, affords an agreeable enjoyment to life. This is the ever-cheerful heart, according to the idea of the virtuous Epicurus. For who should have more reason for being of a cheerful spirit, and not even finding it a duty to put himself in a cheerful frame of mind and make it habitual, than one who is aware of no intentional transgression in himself and is secured against falling into any?*[103]

Not only should our ascetic practices breed virtue, harmonizing our empirical character with our transcendental, but eventually it should (and will) be accompanied by a *positive feeling*. Not merely health, i.e. my virtuous character which makes sure no *ill* maxims will come to action (preventing me from assaulting my brother way before I have even considered its moral reprehensibility) but an *ever-cheerful heart*. To be cheerful is not again itself a demand of reason, as Kant states, but a state of mind that the practitioner of Hellenistic virtuous training will be provided with. If there is anything in the Kantian corpus that inches close to Taylor's wish for a way to understand "loving the good", it is this positive, intuited, *feeling*; the cheerful spirit.

With this we hope to have shown that Kant's philosophizing on virtue is more elaborate that often believed and is capable of dealing with a number of the misgivings that gave rise to the resurgence of

[103] Kant, The Metaphysics of Morals, 6:484-485/p.245

virtue ethics as a discipline. A proper account of love for the good, the importance of friendship (for this see Kant, *Metaphysics of Morals, 6:470-6:473/p.232-235*) moral education (we postpone this issue to section 4.5) and of course, exceedingly important, the issue of having the proper moral motivations for your actions. So, not only do we have reason to attempt a Kantian account of virtue, we can look to Kant's own works in working out this attempt.

Section 2.5: Striving

We have argued for the importance of tying up Kantian morality to an account of virtue due to the problem of a lack of normativity in an account of virtue completely separate from a Kantian account of morality. We have argued that the dual structure of the self we had introduced in chapter one allowed the Kantian to model virtues in a way that they can be understood as normative nevertheless. Finally we have argued that the Kantian program has satisfactory answers to a number of the original problems that induced us to reinvigorate accounts of virtue. Now what is left for us is to sum up where this leaves us in regard to Kantian virtue, looked upon from a distance.

Kantian virtue does not deliver on one thing, the normative autonomy of virtues. Being courageous cannot be understood to be a norm that is put upon me *directly*. If I fail to be courageous this is a lack of my empirical self, something I cannot simply will to be different. Like illness and degraded muscles are lacks of my bodily self, that I cannot be held accountable for in the moment (for I cannot will myself healthy and strong), so also the defects of my mental self (cowardice, dimwittedness) are not my *direct* responsibility. Yet, like we may be held accountable for *not eating healthily* and *not exercising*, which eventually lead to the bodily lacks (and are morally troubling because it leads specifically to vice), so we may be held accountable for shying away from activities that support our mental lives.

This leads to the most general shift in attitude that we find towards virtue in Kant. We cannot be asked to *be* virtuous. This lies outside of the scope of the normative. Yet at the same time, we are always bound to *strive* to be virtuous. In contrast to duty, which should be *done in the moment, irrespective of one's own specific empirical constitution*, Kantian virtue presents us with a picture of morality that is essentially *future-oriented, intimately tied to our empirical character and open-ended*. And we should note

that it makes sense that the future-oriented virtues are tied to our empirical character, for as we have discussed in chapter one, our empirical self is that part of the self which is caught in the forms of intuition, time and space. In contrast, adherence of duty is tied up to our transcendental character and as such is not tied up to any time and place as our transcendental character is outside of the forms of intuition. Adherence to duty is timeless, and can thus only find its expression in a singular moment; the practice of virtue is the expression of duty in the empirical domain; it is duty in time and space and is therefore applied to, and adhered to, by ourselves in time and space. While duty is started, executed and ended in a single moment, virtue is extended over time, stretching forward as far as our own existence. The duty to strive to be good cannot come to an end, for it tells us to always work on ourselves and better ourselves; to rise above the person we were yesterday, to always apply *self-improvement*. Striving to become both *moral* and *capable* becomes the essence of the Kantian view of virtue; a molding of the self towards an ever changing view of the best version of oneself one can be.

Chapter 3: Neo-Kantians on the Self

Deflating the Kantian self and the problems evil and self-regulation

Abstract: In this chapter we will make an excursion to the views of two neo-Kantian philosophers, Christine Korsgaard and Onora O'Neill, whose proposals aim to deflate the Kantian picture of the self. We start by expounding Korsgaard's arguments in *Self-constitution* on where the force of moral commands derives from. We will find that her account of the person's constitution fails however, on account of its incapacity for dealing with *a strong conception of evil*, and in effect, fails a test of normativity. We move on by showing how Kant makes it possible to conceive of a *strongly evil person*. This is done by introducing Kant's views on freedom, and introducing more structure to the nature of the transcendental self. Then we move to O'Neill's account of the person which is characterized as constructed by reason. By taking a pragmatical stance, O'Neill is capable of evading the problems that troubled Korsgaard's account. However, we will argue that O'Neill has problem of making clear in what sense we remain *autonomous*; where Kant has a story on giving the law to ourselves, O'Neill lacks such an account.

Section 3.1: Korsgaard and constitutive principles and defective action

Korsgaard introduces a way to determine specific actions as either moral or not. She claims that adherence to the hypothetical imperative (which she understands as the claim that we must view our actions as effective) and the categorical imperative (which she understands as the claim that we must understand our actions as free) makes an action moral.[104] This is all very much in line with Kant's own views of course. What however is not distinctly Kantian is Korsgaard's insistence on the importance of the idea of a *constitutive principle*. Korsgaard introduces a constitutive *standard* as a standard that:

> *"[…] arises from the very nature of the object or the activity to which it applies. It belongs to the nature of the object or activity that it both ought to meet, and in a sense is trying to meet, that standard."[105]*

---

[104] Korsgaard, The Constitution of Agency, p.9-11
[105] Korsgaard, The Constitution of Agency, p.7

A constitutive *principle* is then a constitutive standard that applies to an activity.[106] These principles are normative and descriptive. Doing the activity exactly means adhering to this principle to some extent. I might fail to adhere somewhat, but at some point my lack of adherence will imply that I'm no longer undertaking the activity. One can swim badly, splashing around in the water failing to make headway, or one might not be swimming at all, never having intended to swim and make headway. [107]

What rational action and agency consist in according to Korsgaard is adherence to the constitutive principles of action in general (the hypothetical and categorical imperative). And it is exactly insofar as we act rationally that we are agents (agents being exactly those things that act 'well').[108]

Constitutive principles of action are both constitutive of *acting well* and constitutive of *being a good agent*. The reason we want such a connection is because we want to make sense of *some person's bad actions* being reflective of *that person's immorality*. They must be responsible for their actions. So, let us look at how Korsgaard argues this:

> *"What does it mean for an action to be defective? The function of a house is to serve as a shelter; a house with a leaky roof tries and fails to be a shelter, and therefore it is a defective house. The function of a sentence is to express a thought; a sentence without a verb tries and fails to express a thought; and therefore it is a defective sentence. The function of an action is to unify its agent, and so to render him the autonomous and efficacious author of his own movements. An unjust or unlawful action therefore fails to unify its agent and so fails to render him the autonomous and efficacious author of what he does."*[109]

For Korsgaard then, defective actions are exactly those actions that do not unify us as agents. However, there is a problem with that this makes it unclear how we should evaluate people who act badly. In how far are they even still agents?

[106] Korsgaard, The Constitution of Agency, p.8
[107] Korsgaard, The Constitution of Agency, p.11
[108] Korsgaard, The Constitution of Agency, p10-13
[109] Korsgaard, Self-Constitution, p.161

Korsgaard's way to deal with this is by making use of the metaphor of the state as standing for an agent (an idea she extracts from Plato's *Republic*)[110]. Defective action is just like having a *defective constitution*. Some state's constitutions are better at unifying the state, and some constitutions are worse. In the same fashion in functions within the agent. When we have, as agents, a bad constitution, we are still acting in accordance with our constitution, but we are acting on a constitution that is simply worse at unifying ourselves. This means that acting badly is not binary, we do not act good or bad, we act more or less defectively. There are many possible constitutions that are defective, the one worse than the other.

Section 3.2: Determinedly evil and the capacity to fail

Can there be a normative standard that is impossible adhere to?[111] We expect people to be moral. One *ought* to be moral, and the bar that we set with our normative demands for morality is a bar that places a *demand* upon us. Of course one can doubt[112] there might be something like this, but insofar as reason is of interest, it should incorporate this feature. Kant famously introduced that:

---

[110] Korsgaard, Self-Constitution, p.133

[111] It is in the affirmative answer to this question that we find the most potent formulation of naturalism, and in it, its greatest threat. It is this naturalist threat that Kant himself famously anticipates in the critique of pure reason when he writes: "Thus I had to deny knowledge in order to make room for faith" (Kant, Critique of Pure Reason, BXXX/p.117). If we are to be expected to adhere to moral commands for example, we should expect us to be able to do so. But, given that we are biological creatures, limited by our bodies, upbringing, social status, evolutionary tendencies, etc. we cannot expect everyone to adhere to this. Perhaps one's upbringing was so strictly conservative that his defense of racist remarks cannot have been avoided by himself. Connecting this to our insistence on the possibility of being reasonable, we are left with a severely troubled account of reason. We believe that this is the core of the naturalist position.
Kant's genius move was to show how it is in fact possible to understand ourselves as free in the noumenal world (and in that sense capable of doing both good and bad) all the while accepting the naturalist observation that we humans are also causal natural entities.

[112] Or at least discuss; most authors who argue in favor for one or another bar set by reason would probably claim that in the end, when all is understood well, we cannot truly doubt its existence. Nagel's reading of Descartes puts doubt of the cogito on the level of a contradiction (The Last Word, p.19). And as Anselm argues, we may entertain contradictory statements through words, but not through the objects signified, that is to say, *we cannot properly think a contradiction* (Anselm, Proslogion, p.88-89).

"Now of course the action must be possible under natural conditions if the ought is directed to it"[113]

For, how can we expect the impossible? We would like to put it like this; with an impossible demand we lose the normative character of *the demanding*; an impossibly high bar set by reason does not make everyone unreasonable.

Next we can of course wonder: can we then not still have a bar of normativity that is set so low that it becomes impossible to fail when we do our act (of thought)? But this also cannot be made sensible. What we identified as essential to any interesting account of normativity is that there is some standard, adherence to which is preferable and failure to adhere to deplorable. However, if the bar is so low that adherence is guaranteed, how can this then be presented as a demand on someone? Nor does is seems sensible to commend someone for what is necessarily so. Fichte put the point (somewhat convolutedly) like this:

"Opposition in general is posited absolutely by the self."[114]

Or as Pinkard, in his discussion of Fichte, elaborates:

"In a rather dense and compressed series of arguments, Fichte concluded something like the following. To adopt any kind of normative stance at all is to commit oneself necessarily to the possibility of negation., of asserting not-A."[115]

Fichte argues that, the only way that we can understand a normative demand on us (for example on our beliefs), is by knowing that they can be both true and false. For if our beliefs could only be true, there can

---

[113] Kant, Critique of Pure Reason, p.540
[114] Fichte, The Science of Knowledge, p.103
[115] Pinkard, German Philosophy 1760-1860, p.115

be no standard against which we can decide them to be the sort of beliefs we would prefer. So, we require the idea of the 'not', the idea of untruth, to make our true beliefs *valuable.* Normativity implies the necessity of the possibility of being wrong.

This means that before some demand is normative on a person's actions, we need that person's actions to potentially not adhere to the demand (and of course also potentially adhere). If either is impossible, we have shown that the demand is not properly normative. This, combined with what we have discussed about Korsgaard, will allow us to deliver a potent criticism of Korsgaard.

Firstly, the constitutive principles of action in general are also constitutive of being an agent. But, what then happens when we act badly, when we do not adhere to the demands of reason? When we do this we cease to act at all and through that cease to be an agent. If this is the case however, can we even conceive of an agent not adhering to the demand of practical reason? We cannot, any agent by definition already adheres to the demand of practical reason. But then, it becomes impossible to fail (as an agent) this demand of practical reason. If this is so, as we have discussed before, it makes no sense then to see this as a proper normative demand. A constitutive principle for actions and agenthood cannot be normative for an agent, for an agent cannot fail to adhere to the principle, for when he does not adhere, he is no more.

In a way the constitutive principle parcels out the ability to fail and succeed at reason out to two entities. The reasonable creature is exactly the one that can adhere to the principles of reason, and the unreasonable creature is exactly the one that is capable of failing to adhere to the principle of reason. The problem is of course is that we want *one and the same person* to be capable of succeeding of failing at reason, because we want the principles to be normative *for him*.

As we have seen however, Korsgaard has a slightly more elaborate view on what she calls defective action. Here, we are proposed that not adhering to the demands of reason is not binary, one can be more or less unreasonable, and that what makes this possible is our 'constitution'. We are still unified into one person, but less capably. On this view, our actions can be unreasonable while we remain to be agents. However, this move does not truly solve the problem, it only pushes it back. To some extent, we can now understand how we can fail specific principles in our concrete actions. At the same time, what we now can ask how it can happen that we fail at being reasonable by incorporating the incorrect constitution. We are left with the same problem; we cannot fail at having the correct constitution, because that constitution defines our identity; when we fail at having the best constitution, we simply are this

other, lesser constitution. Even with the introduction of slightly better and worse actions and constitutions, we are still left with that these are constitutive for agency. Let us follow Korsgaard's metaphor of the state: there is a large distinction between states with different constitutions. The Dutch Republic and the Kingdom of the Netherlands show some continuity, but are also clearly different things. Their constitutions make them into the states that they are and having differing constitution makes it difficult to identify the one as the other.

Even if we leave this issue of partial identity aside, Korsgaard's solution remains troubled. For, even though we are presented with a sliding scale of constitutions, Korsgaard is still not able to present us with the constitution we would truly require, one that (in the moral domain) is truly *evil*. That is to say, Korsgaard introduces how there may be partially unreasonable agents, but does not deliver an explanation to our original point; how can we (properly) fail at being reasonable while retaining our agenthood. We repeat ourselves; we need such a possibility for the demand of reason to actually be normative. We need, as we will call it, the possibility of a *determinedly evil person*. Korsgaard on the possibility of such a person is however ambiguous:

> *"[…] the tyrannical person* ***doesn't choose an act for the sake of an end****, the whole package as something worth doing. There's one end – as in the case of the serial killer, it may be the act itself – one end or act he's going to pursue or to do no matter what, and it rules him. And for him that end makes anything worth doing, anything at all, and that's a fact that is* ***settled in advance of reflection****"*[116] [emphasis ours]

The tyrant is not to be understood from the outside, he is maniacal, and in the end, not an actual agent. He does not undertake actual actions (acts combined with ends) nor does he reflect. He has ceased to be a rational being by Korsgaard's own standards. But, we require the tyrant to be in fact an agent. We need this not because of what troubles Korsgaard here, where she defends herself against a possible absurd effect of her theory:

[116] Korsgaard, Self-Constitution, p.172

> *"But if tyrants really are tyrannized over by some force within them, whose thing they have become, are they responsible for their condition? And, more importantly, are we to say that the other characters become less responsible to the extent that there is less of agency about them, and more of the operation of some external or internal force?"*[117]

Her answer is no, excluding perhaps the fully tyrannical "person", not acting does not exclude one from moral reproach. However, this was not what we took issue with. We don't need to be able to reproach the lesser agents and the tyrant, we needed the tyrant to be viewed as coherent because we need an example of a *determinedly evil person*. If we cannot conceptualize such a thing, failure to adhere to reasons demands remains an impossibility, and the normative character of the demand becomes void. The impossibility of evil action disturbs the normative character of good action.

Section 3.3: Freedom and evil

In the previous sections we have seen how Korsgaard's attempt at providing us with an account of practical reason failed due to the inability to account for the determinedly evil person; agency and (good) acting were so closely intertwined that we could not conceive of an agent who is truly evil. However, this may prove illuminating for other attempts at providing an account of practical reason; how do authors like Schelling, Kant and O'Neill provide us with ways to deal with the determinedly evil person?

The discussion will center around the following group of concepts: the person (or self), the (morally good) action and the freedom of the person to choose actions. We have already seen the importance of investigating the relation between the person and her actions; when they become too intertwined there remains no possibility for a (proper) person to act badly at all. But why do we include freedom? We do this because *freedom is a characterization of the relation between some person and her actions*. We can rephrase Korsgaard's problem in terms of freedom; are we free to choose the bad action? The answer is *no*, Korsgaard's account introduces a modal impossibility for *my* acting badly (due to my

---

[117] Korsgaard, Self-Constitution, p.174

ceasing to be *me*, or any person were I to act bad) which means I'm not free to choose to act badly. And this reflects poorly on the question whether I am free to act properly; for given that bad action is impossible, we cannot state that we are free to *not* act badly, i.e. well.

By introducing the concept of freedom in this context we draw upon more literature to look for solutions to the problem. Specifically, Kant's and Schelling's attempted solutions revolve around the attribution of very strong conceptions of freedom to persons. O'Neill explicitly distances herself from the strong conception of freedom as found in Kant and proposes a more pragmatically inspired account of both freedom and the person. All these might function as alternatives to the conception of the self that lead to problems in Korsgaard.

Section 3.4: Intermezzo through Schelling; symmetry and asymmetry

Before we move to Kant's and O'Neill's specific solutions, let us take detour through Schelling's views on evil to provide more general terminology that will help us evaluate Kant and O'Neill. In the *Freiheitsschrift* Schelling provides us with an investigation into the notion of freedom; and with it an investigation into what sort of ontology is required for freedom to be able to exist. Schelling states that:

> *[…] the real and vital concept is that freedom is the capacity for good and evil.*[118]

Freedom is the possibility to do *both* what is good and what is evil. We will call this a *requirement of symmetry*; for freedom to exist, there needs to be some symmetry between evil and good, for they both need to be possible. That is to say, a free person is exactly a person who is free to choose to either act badly or to act well.

[118] Schelling, Investigations into the Essence of Human Freedom, p.23

It is this requirement of symmetry that was not met in Korsgaard's account of practical reason. At the same time, it was exactly this lack of symmetry that she leveraged to provide us a vindication of her moral principles. The immoral could be identified with the impossible. Immoral actions are those defective actions that are performed by *non-persons*. The moral actions are identified with the possible (or necessary). Moral actions are those actions which are performed by *persons that are unified by their action.* Korsgaard in this fashion could show us in what way the moral action differed from the immoral. So we find that not only there must be a symmetry between moral and immoral action, there also needs to be a way to decide *which of the two actions is the moral (commendable) one and which of the two is the immoral (reprehensible) one.* This is *a requirement of asymmetry* on an account of practical reason.

We can see in Schelling's further remarks that his struggle is with exactly the tension that arises between the requirement of symmetry and the requirement of asymmetry. As we saw, he starts with defining freedom in a way that puts evil and good on equal footing. Schelling asserts that for freedom to be possible we need to make sense of the substantial existence of both a principle of good and a principle of evil.[119] If we don't, he asserts, we make it impossible to make a substantial choice for either principle. Freedom then consists exactly in being free to adhere to either the evil or the good principle; a choice for evil cannot be *only* our lack of choice at all. But this means for Schelling that we need to 'share' in both ontological principles. What's more, this is exactly the specificity of human nature, to share in equal parts in the principle of the good and the principle of evil:

> *This raising of the deepest centrum into light occurs in none of the creatures visible to us other than man. In man there is the whole power of the dark principle and at the same time the whole strength of the light. […] The human will is the seed—hidden in eternal yearning—of the God who is present still in the ground only; it is the divine panorama of life, locked up within the depths, which God beheld as he fashioned the will to nature.*[120]

But now Schelling finds himself in a predicament, how can we, if the world contains two principles, a good one and an evil one, decide which one is the good one and which one is the evil one? What are the

---

[119] Schelling, Investigations into the Essence of Human Freedom, p.23-24
[120] Schelling, Investigations into the Essence of Human Freedom, p.32

differences between these principles that allow for the (moral) asymmetry? Schelling's answers in good pantheist fashion[121]; God, the totality in which all things are immanent, is the principle of the good. But in stark contrast to other pantheist authors, Schelling has decided that beside the principle of the good there should be a principle of evil which cannot be fully characterized in terms of the (lack of the) good. So, for Schelling, evil cannot consist in the absence of good, as is a usual strategy, but consists in *resistance* to the principle of the good.

But now we approach an apparent contradiction; if the good is supposed to be that in which everything is immanent, how can there be a separated principle of evil with its own independent existence? The answer lies in that things can retain their own separate existence, their own essence, even when contained within, or being dependent upon, a larger whole, he remarks:

> *Every organic individual exists, as something that has become, only through another, and in this respect is dependent according to its becoming but by no means according to its Being.*[122]

The principle of evil then takes on a similar guise; although evil is contained within the principle of good, it has an essence that is independent from the good. Evil is exactly that power which resists the good; in other terms Schelling expresses it as the resistance of (evil) particularity against (good) unity.

Schelling pushes forward; the work done by unification can only attain any dynamism when it has the resistance posed by particularity; both principles exists in mutual co-dependence. But this co-dependence is not symmetrical; *unity works on particularity* and thus carries particularity within it,

---

[121] Schelling explicitly endorses *some* form of pantheism, and as he puts it, deciding exactly how we should understand pantheism is the issue at large (Schelling, Investigations into the Essence of Human Freedom, p.11). Spinoza is both an opponent (due to Spinoza's fatalism) but also an inspiration for Schelling here (Schelling, Investigations into the Essence of Human Freedom, p.12). Another source of inspiration for Schelling's thought is as he calls it "Platonic matter" (Schelling, Investigations into the Essence of Human Freedom, p.41). We take Schelling to refer to Plato's dialogue *Timeaus*; here we find God as a craftsman (*demiurge)* who brings unity into existence. But while doing so, the *demiurge* is restricted in all kinds of ways; both the form of the universe but also the *particular* nature of matter *resist* the work and wishes of the *demiurge* (Plato, Timaeus). In a similar fashion, the principle of evil *resists* in its particularity the unification the principle of the good attempts to press upon it. Important for Schelling is to notice that since the matter can resist the *demiurge* we need to accept that the matter has an *existence for its own*.

[122] Schelling, Investigations into the Essence of Human Freedom, p.17

particularity resists unity and is the *ground for the functioning of unity*. The symmetry exists in the essentially dual structure of Schelling's ontology, both principles have their own distinct existence; these cannot be reduced to each other. At the same time the two principles differ, the good is identified with the totality of God's being, evil with the necessary *particularity* that *grounds* any application of God's unitary motive.

As we can see, Schelling's approach is based on heavy metaphysical machinery. In an attempt to understand freedom, Schelling is compelled to put good and evil 'into the world'. And we see that Schelling remains troublingly metaphorical. His construction however was made in the awareness of the importance of two properties of freedom that need to be accounted for; the symmetry between the possibility for good and evil and the asymmetry needed to prefer good over evil. What he also understood is that understanding evil as simply the absence of the good is a worrying strategy. When doing so, evil becomes fully derivative. But what remains then of our capacity for evil? We can no longer understand ourselves as freely choosing for evil; evil then turns into a *weakness of the will*. The will is then understood as incapable of pursuing its telos, its capacity for good. Schelling is on the right track when he asserts that freedom and a derivative notion of evil are in tension with each other. We will see Kant and O'Neill work in a similar vein.

Section 3.5: Kant: between *Wille* and *Willkür*

For Korsgaard one's action and one's personhood cannot be torn apart from each other. For Kant this is different, the relevant properties required for personhood are not connected with actual actions, they are connected to *our potency for action*. This means that we want an analysis of which things are prerequisites for acting, for if some creature shares in these prerequisites, we can claim that it is in fact a person that could be held accountable for actions. This is all not connected to a person's *de facto* acting or not-acting, the potency for action is not through that destroyed or constituted. Of course, Kant is well aware that there is no way for us to ascertain that another creature is a potential actor if he never acts at

all. Despite that, this potentiality is an essential part of the theoretical construct, required for the vindication of the Categorical Imperative as the supreme principle of practical reason.[123]

This means for Kant that there has to be some fact of the matter regarding the status of being a person; some substrate that remains when the person is stripped from all *acting*. This highly abstract *transcendental self*, is characterized only by those properties that are general to all those things that are *selves*. Manfred Kuehn in his biography of Kant puts it nicely when he contrasts Kant with Hume:

> *Hume thought he could account for moral judgement in terms of a "pleasing sentiment of approbation" by an unbiased and disinterested spectator. Kant develops the idea of a completely rational observer of himself, or perhaps better, of an agent split in two, namely, a non-rational actor and a rational observer of these actions.*[124]

All people who have the potency for moral action need to be of this 'split' nature. Everyone has a distance to themselves that is prerequisite for evaluating one's owns maxims. Preempting the idea of *autonomy* we will discuss further on, we should say that *if we are to give ourselves the law and thus be autonomous* we need to have a constitution that allows for this being possible.

However, Kant goes further than to claim *just* that we should and can reflect on ourselves and our maxims. In the practical domain, there is something we can assert of this *transcendental onlooker*, something that on Kant's view, is another prerequisite for any moral action at all. This property of our

---

[123] We see this idea return in all of Kant's ethical writings. In the *Critique of Practical Reason* as we will see, we are told by Kant that being free is one of these prerequisites. Were we not free, our acting on *bad* maxims would not be something that we could be held accountable for. In *The Metaphysics of Morals*, one of two conceptions of virtue revolves around the idea of *making ourselves the sorts of beings that can act morally*. In the *Groundwork for the Metaphysics of Morals*, Kant identifies a property of the person as the morally relevant element: the goodness of someone's *Wille*. Even when not acting, the good *Wille* retains its moral character. And finally in *Religion Within the Bounds of Reason*, Kant posits two *principles* that inhere in human nature (pre-empting Schelling's ideas), a good one and an evil one. It is this specific character of human nature that lays open to us the possibility of (im)moral action. In all cases, Kant seems to believe that the specific constitution of the person is of essence in the analysis of what the highest principle of morality is.

[124] Manfred Kuehn, Kant; A Biography, p.202

self-observance can in no way be intuited and thus lies beyond the realm of knowledge as was delineated by Kant in the Critique of Pure Reason.[125] This property is the self's *freedom*. Kant states:

> *It is therefore the moral law, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves [so bald wir uns Maximes des Willens entwerfen]), that first offers itself to us and, inasmuch as reason presents it as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them, leads directly to the concept of freedom.*[126]

We should be careful in looking what exactly is attributed freedom on Kant's view. The 'free will' can have in the Kantian architecture three very distinct meanings.[127] This is because for Kant 'the will 'as we commonly use the term is ambiguous. On the one hand there is *"Willkür"* (which is sometimes translated as will, sometimes as choice) and on the other "*Wille"* (always translated as will). *Willkür* is meant to designate our capacity for coming to action, for choosing between different maxims available to us; it is our capacity for choice. *Willkür* thus designates the ability to *decide* which ends we will act on and which means we will apply to do so. Thirdly (we will return to this at the end of the section) there is the conjunction of *Wille* and *Willkür* (also termed *Wille* by Kant) which can in its own way also be free.

---

[125] In the Paralogisms of Pure Reason, Kant discusses what it is that we can and cannot conclude from the fact that we can know that the *I* in the *I think* exists. He charges rational psychology, most notably exemplified by Descartes' work, with concluding far to much. Derivations that result in the conclusion that the nature of the *I* is a soul (or thinking substance), or about the *I's* eternal existence (Kant, Critique of Pure Reason, B406-B413/p.415-458) are analyzed and ultimately cast aside. Longuenesse in her work on the self in Kant's philosophy *I, Me, Mine*, discovered an interesting parallel between Kant's paralogisms and his lecture notes on metaphysics from the late 1770's. Here we find Kant make derivations of exactly the kind that he in the Critique of Pure Reason was to refute and identify as paralogisms. The first four paralogisms and his four first derivations pair up nicely. However, in the lecture notes, Kant provides a fifth derivation in a similar fashion, one of the spontaneity of the self. Longuenesse points out that spontaneity is conspicuously absent in the paralogisms of pure reason. (Longuenesse, I, Me, Mine, p.154-155) Why did Kant leave out the analysis of this derivation as a paralogism? The answer lies in the fact that Kant, as we see most explicitly in the Critique of Practical Reason (Kant, Critique of Practical Reason, 5:29-5:30/p.27) does in fact subscribe to the spontaneity of the self. Even in the Critique of Pure Reason itself, Kant already characterizes pure thought as *spontaneity* (Kant, Critique of Pure Reason, B132/p.246). Longuenesse herself proposes to analyze the attribution of spontaneity as a paralogism as well (Longunesse, I, Me, Mine, p.155-158).

[126] Kant, Critique of Practical Reason, 5:29-5:30/p.27

[127] Allison, Kant's Theory of Freedom, p.129

Wille is a subtle capacity of our thought. Where *Willkür* is the possibility to choose between maxims, *Wille* is the capacity to prefer maxims on the basis of their form; and in effect "provide the norm"[128] for the *Willkür* . Why should this specific way of preferring maxims (on their form and not their ends or means) be sharply delineated from preferring maxims on the basis of other characteristics than their form? The reason for this is that preferring maxims based on their form is the only way of preferring one maxim over another that is not necessitated by that which is external to the faculty evaluating the maxim itself. A faculty that evaluates the form of the maxim can be called *independent*.[129] That is to say, the *Wille* is exactly that capacity of our practical reasoning that *simply is* independent of *external necessitation*. When we prefer a maxim based on its form we are *not induced to a specific choice by any elements foreign to ourselves*. We are not independent when we consider anything else about the maxim; either the ends or the means (whether we want a specific end depends on our contingent desires, whether we want to apply certain means depends on the specific and contingent way the world is). Kant states in the *Critique of Practical Reason*:

> *If a rational being is to think of his maxims as practical universal laws, he can think of them only a principle that contain the determining ground of the will not by their matter but only by their form.*[130]

Wille then is essential to the Kantian moral project. It is our capacity for considering the form of our maxims, and as such, allows for the application of the Categorical Imperative (a demand on the forms of our maxims) to specific maxims that we consider acting upon. To fail to heed the insight of the *Wille*, the insight into the properness of the form of our maxim, is exactly what it means to be heteronomous, and to have acted badly.

The *Willkür* meanwhile, is not independent. *Willkür* needs to be induced by goals that are wished for by the person and means that are rationally derived from knowledge from the outside world. Both goals, and means for them, rob the *Willkür* of its independence. It cannot come to its own decision, it is

[128] Allison, Kant's theory of freedom, p.130
[129] Allison, Kant's theory of freedom, p.131
[130] Kant, Critique of Practical Reason, 5:48/p.41

induced to let the content of the maxims it is considering determine itself. *Willkür* can only prefer maxims on the basis of considerations external to the activity of *the consideration of maxims an sich*.

To state of the *Wille* that it is free, over and above independent, does not add anything to our understanding of *Wille*; it is not the capacity for choice and as such we can expect no more than independence from this faculty. Nor is the existence of *Wille* according to Kant anything spectacular, we can all test for ourselves whether we are capable of examining the form of a maxim. That is to say, *Wille* is a property of ourselves that we *can* intuit.[131]

But then where is this supposed concept of freedom that lies beyond all possible experience? This freedom is ascribed of *Willkür*. We are free to *choose*, that is to say, we cannot be described as being completely determined by the desirability of our ends. Most importantly, this allows us to also *consider* the judgement given to us by the *Wille*. The *Wille* is that which can be motivated by the form of the maxim. And it is clear that at times we are motivated by the *Wille*, that is to say, we choose through *Willkür* a maxim on the basis of the motivation of the *Wille*. But, sometimes we do not do so as well. Sometimes we are motivated solely by our ends and desires. In those cases *Willkür* does not consider the judgement of the *Wille*. Now, the importance of freedom of the *Willkür* is to *make certain* that at any moment we *could* have let the *Willkür* been determined by the judgement of the *Wille*. It guarantees our *potency for both good and evil actions*.

---

[131] We do at times consider the form of our maxim. This we can observe ourselves doing. So, we can also observe our capacity for evaluating (at times) our own maxims. Kant believes this dual structure of *non-rational actor* and *rational observer* folded into one person, is something we can indeed be aware of. When we experience that we can reflect, we experience ourselves as being split up, being both thinker and object of thought. This not only occurs in his ethical writings, in the Critique of Pure Reason, the transcendental apperception incurs a similar idea. Here too, we can be aware of our split nature, of our noumenal *thinking* existence, which is known through thinking it (but only known in its unifying power, its existence as thinking and nothing more), and the way this phenomenally is presented to us through inner sense i.e. the form of time:

> *"In the transcendental synthesis of the manifold of representations in general, on the contrary, hence in the synthetic original unity of apperception, I am conscious of myself not as I appear to myself, nor as I am in myself, but only that I am. This representation is a thinking, not an intuiting. […] I exist as an intelligence that is merely conscious of its faculty for combination but which, in regard to the manifold that it is to combine, is subject to a limiting condition that it calls inner sense, which can make that combination intuitable only in accordance with temporal relations […]"* (Kant, Critique of Pure Reason, p.259-260/B157-159)

Or as Longuenesse nicely calls it, there is in Kant *a mere consciousness of the act of thinking* which points toward a split that exists between the un-empirical *action-awareness* and the empirical *object which acts.* (Longuenesse, I, Me, Mine, p.86-87)

As alluded to, there is a third sense of freedom of the will in Kant. *Wille* can be understood both narrowly and broadly.[132] We have discussed the narrow *Wille*. The broad *Wille* is the conjunction of narrow *Wille* and *Willkür*; that which we should properly call practical reasoning. This broad *Wille* is free in the sense of being autonomous; it can give itself the law to follow. This freedom can only occur as an effect of both the freedom of *Willkür* and the freedom of narrow *Wille*. The narrow *Wille* provides the *Willkür* with the law (through its independence from alien authority), the *Willkür* can decide (freely) to follow this law. Lawgiving is done by narrow *Wille*, and law following by *Willkür*, it is their conjunction that can be said to be *giving the law unto itself*; *Wille* broad, or practical reason, can then be called *autonomous*.

With this we have discussed three ways in which we are free. We have (and can intuit) an independent faculty which is the narrow *Wille*, which is motivated by the form of a considered maxim. We have the unintuitable noumenal freedom that is ascribed of *Willkür*; our freedom of choice. And thirdly we have the derivative notion of freedom, ascribed to the conjunction of the narrow *Wille* and *Willkür*. This is our autonomy, i.e. our capacity to give the law unto ourselves. When we are autonomous we can be said to be good; in that case we freely chose to give the law to ourselves.[133]

Let us return to our original problem: how does Kant deal with the evil person? How does he make sure that the choice between good and evil maxims is symmetrical with regard to possibility? The answer lies in the assertion of the freedom of the *Willkür*, because the *Willkür* is capable of choosing between being motivated by the *Wille* and thus the form of the maxim, or by the content of maxims. Both are asserted as always being possible because freedom is asserted of *Willkür*, and this freedom is inseparable from this *Willkür*; no actions (from their evilness alone) can leave the *Willkür* unfree. But how then can we still distinguish the good maxim from the bad maxim? What allows for asymmetry? This is of course the dual structure of *Wille* and *Willkür*. Our freedom of *Willkür* ensures our demand of symmetry, while the norm put forward through the independence of the *Wille* ensures our capacity for deciding what maxims are acceptable and which not. Bad maxims are followed by those who do not heed the judgement of the narrow *Wille*, and who are in effect not autonomous. So, when we act wrongly, we do not sacrifice our agenthood, as in Korsgaard's case, what we sacrifice is our *autonomy*. We choose freely to be

---

[132] We borrow this terminology from Allison's *Kant's theory of freedom*. Kant himself never explicitly states the difference in these terms himself. (Allison, Ch7 Wille, Willkür and Gesinnung, p. 129-145)
[133] Kant, Groundworks for the Metaphysics of Morals, 4:440-441/p.58-59. This is one of the places where autonomy of the will is explicitly introduced as the supreme principle of morality and heteronomy as that of immorality.

determined from the outside, to not heed the judgement of the *Wille*, to not be autonomous. It is our lack of autonomy that coincides with our bad action.

We see this point return and made more sharp in Kant's *Religion Within the Bounds of Reason Alone.* There are two overarching types of maxims; the principle of self-love and the principle of justice. We are at all times free to adhere to one of these two principles. What's more, we at all times have *inclinations* pulling us in either direction. There is in the very nature of a person both a tendency for, and freedom to do, evil (i.e. self-love) and a tendency for, and freedom to do, good (i.e. adhere to justice). These two principles both have individual existence, where the one can exist without the other.

The angel, or *holy will*, is on Kant's view exactly a creature that lacks any empirical self, and as such, there is no possibility for the angel to ever act on its principle of self-love (there is no self to love). However, the angel or holy will, not being in any way disturbed by outside influences and without desires, were it to consider a maxim, could *only* consider the maxims form:

> *A perfectly good will would thus stand just as much under objective laws (of the good), but it would not be possible to represent it as necessitated by them to lawful actions, because of itself, in accordance with its subjective constitution, it can be determined only through the representation of the good. Hence for the divine will, and in general for a holy will, no imperatives are valid; the ought is out of place here, because the volition is of itself already necessarily in harmony with the law. Hence imperatives are only formulas expressing the relation of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, e.g., to the human being.*[134]

In such a creature the *Wille*, and the principle of justice, would reign supreme, without any resistance provided by the empirical side of the self. Kant remarks that such a creature cannot be said to *adhere to an imperative*, for he necessarily acts in accordance. What's more, such a creature does not possess the free *Willkür*; there is in the end only one possibility and so no freedom to choose.

[134] Kant, Groundworks for the Metaphysics of Morals, 4:414/p.30-31

Notice the differences between Kant and Korsgaard. For Korsgaard, just like for Kant, *autonomy* is the property that imbues our account of practical reason with its normativity; we should be autonomous and we should act on maxims that exhibit this autonomy. Korsgaard however does not introduce *freedom* as distinct from *autonomy*; the freedom of choice of a person is constituted together with her autonomy in good action. There is no room for a freedom of the *Willkür* that lies behind this constituted person, that ties her together: Korsgaard wants all unity of the person to derive from action and autonomy. Kant's introduction of *freedom* as distinct from *autonomy* and the attribution of *freedom of choice* to the person, whatever actions may come, allows Kant to provide a unity of the person *before* the person acts. If we want to be able to hold *people* accountable for *their* actions, we need to presuppose some metaphysical substrate characterized by freedom, Kant argued.

Section 3.6: O'Neill: constructing reason and prying apart persons, freedom and actions

It is not hard to see what is unattractive about the picture sketched by Kant. Freedom lies beyond the realm of experience, and yet we need to assume that it is true of every creature that is characterized by reason. This is one of the few truly *metaphysical* claims of Kant's philosophy. Note the difference with the postulates of practical reason. The eternally existing soul also lies beyond all possible experience; and thus Kant wants us to use it as a postulate in our practical reasoning, while reserving theoretical judgement on the matter. We can act *as if* these postulates hold, but we do not need to truly *assert* their truth. This does not hold for the freedom of the *Willkür*, we *should* assert its truth about our own person. Freedom of the *Willkür* then is true metaphysical *knowledge* on Kant's view. It is then of course very sensible to consider whether we could get Kant's result without this metaphysical crutch.

Many commentators on Kant in one way or another take 'naturalizing' Kant (without loss of normativity) to be an important mission. Resscher takes a demystifying approach to the postulates of practical reason and the *ding an sich*.[135] Korsgaard can be understood as exactly an attempt to work without any 'free-person-substrate' that lies behind our actions.[136] Recently, Longuenesse has taken an

---

[135] Resscher, Kant and the reach of Reason
[136] Korsgaard, Self-Constitution

approach where the similarities between Freud and Kant could be leveraged to garner an understanding of Kant's general conceptions of the self in Freudian, naturalist, terms.[137] Similarly, in the domain of practical reasoning, O'Neill attempts to deflate the concept of the *person*. What the scope of practical reasoning should encompass, and what exactly should be its *locus*, needs to be constructed in an act of reason instead of "already being out there" as with Kant.

O'Neill is well aware that the Kantian picture includes some unpleasant metaphysical elements as she states in the preface of *Constructions of Reason; Exploration of Kant's Practical Philosophy*:

> *Kant is revered for his unswerving defense of human freedom and respect for persons, and for his insistence that reason can guide action. He is also reviled for giving a metaphysically preposterous account of the basis of freedom and an intermittently repellent and simultaneously vapid account of human obligations. Many contemporary proponents of "Kantian" ethics want the nicer bits of his ethical conclusions without the metaphysical troubles.*[138]

In the work cited, O'Neill does not attempt to deflate the metaphysical troubles herself, remaining relatively faithful to the Kantian texts. However, O'Neill does provide us with more of her own altered Kantianism in her entry into the *Oxford Handbook of Rationality* titled *Kant: Rationality as Practical Reason* and her book *Towards justice and virtue: A constructive account of practical reasoning.* Much of her alteration resides in the deflation of the concept of the self. Or, more specifically, the deflation of the importance of the *exact* ontological status of the self.

It is of paramount importance to O'Neill that we see that Kantian autonomy is not something we should ascribe to *persons* or *selves*, but to *reason, ethics, principles* and *willing*.[139] O'Neill's contention rings true if we consider the way autonomy is understood in the Kantian framework; *not being subjected to foreign authority* (autonomy in the negative sense) and *giving oneself the law* (autonomy in its positive sense). In reasoning and willing we may conceivably *give the law to ourselves* and our ethics and our

---

[137] Longuenesse, I, Me, Mine
[138] O'Neill, Constructions of Reason: Exploration of Kant's Practical Philosophy, p.i
[139] O'Neill, Kant: Rationality as Practical Reason, p.107

principles are the right sort of ethics and principles when *they are not formulated subject to foreign authorities.*

The self here gets deflated when we compare it to the orthodox Kantian picture we described before; no longer do we concern ourselves with the person's *Willkür*. The *Willkür* and its freedom provide for Kant a substrate of the morally relevant person, that lies behind our concrete (heteronomous or autonomous) acting. For O'Neill, the relevance of this substrate is limited, as it is autonomy that points us to the good and the bad action, we should not require more than our ability to *recognize* autonomy (the capacity signified by Kant's *Wille*).

This leads O'Neill to make her second move. O'Neill understands herself as having two major problems to deal with when thinking about Kantian ethics:

> I have often thought of these two lines of inquiry as exploring the arguments that lie respectively upstream and downstream from the topics discussed in Acting on Principle. Upstream issues includes Kant's spectacular claim that a vindication of reason is possible; downstream issues include consideration of ways in which normative standards, among them principles of duty, can inform practical judgement and be used to shape action.[140]

The first step in deflating the self was to argue that in an account of practical reason, what we try to get at is an idea of *autonomy*. Then O'Neill argues that in a vindication of *autonomous practical reason* we do not need to discuss the issues of persons or personhood; these problems are to be relegated *downstream*. That is to say, persons are important for Kantian ethics because we need to get a grip on the idea of *the scope of practical reasoning*. The second step is then, according to O'Neill, to tell a story about the applicability of autonomous action to others: this she provides in *Towards justice and virtue*.

One of the worries O'Neill had when the Categorical Imperative tells us to, for example, not use others as merely a means to an end, is how we should construe these *others*. Historically, a lot of major injustices were exactly an effect of construing other groups as *inhuman* and, through that, excluding these groups from any moral consideration at all. We may be capable of following the Categorical Imperative

[140] O'Neill, Acting on Principle, p. 29-30

while also subscribing to some forms of oppression or slavery as long as we claim that these creatures being oppressed are simply not the sort of creatures that fall within the scope of the Categorical Imperative.[141]

So, if there is no metaphysical "fact of the matter" regarding the status as a person as morally (ir)relevant, how does O'Neill propose to construct the scope of moral considerations? Through our implicit acceptance of the capacity for practical reasoning of the other person. If in our own relation to the other we implicitly make use, or presuppose, the capacity for practical reasoning on their behalf, we are also to accept them as relevant (others) for ethical consideration. The idea behind this is that the binding force of the Categorical Imperative is that we understand *our* maxim for doing some action to be *lawlike*. Among other things, this means that our maxim should be *capable* of being subscribed to by any other creature who is capable of lawlike reasoning. So, when we construct someone else as being capable of lawlike reasoning, we also accept that she is the sort of creature that *in principle* needs to be able to subscribe to (or at least, not necessarily be opposed to) my maxim. This leads to the demand to take the person along in ethical consideration; not doing so would make your maxim unlawlike as it has now become in principle un-universalizable to some of those to whom it should be universalizable (others capable of lawlike reasoning).

And for O'Neill, the way to determine who these are (a practical procedure) is by sincere consideration of what we ourselves implicitly assume about those others in our own acting. For example, O'Neill considers the possibility of excluding children and wives from ethical consideration as they might be considered as being dependent on the patriarch of the family and in that sense not being distinct from the patriarch. O'Neill argues against this possible excess of not pinning down ethical standing metaphysically:

> *Allegations that 'dependents' are not distinct beings spuriously invoke fictions of ideal integration which do not in fact inform attitudes, feelings or involvements. In practice, most activity directed at children, wives or other supposed or actual 'dependents' acknowledges and assumes that, on the contrary, they are distinct agents and subjects, with multiple, rather useful capacities and capabilities to act and to feel, to respond and even to take initiative. […] If separateness and*

[141] O'Neill, Towards Justice and Virtue, p.91

> *plurality are assumed in activity it is inconsistent to deny them and to substitute ideal conceptions of integration and unity in considering the appropriate scope of consideration.*[142]

In similar fashion she argues against the possibility of slavery and other unjust/inconsistent distributions of ethical standing that occur and have occurred. O'Neill's hope is of course that by providing a practical procedure for deciding who we should attribute ethical standing we can forego the metaphysical discussion on who or what *exactly* should count as the being that is relevant for moral consideration.

In contrast with both Kant and Korsgaard, O'Neill provides us with an account of practical reason that makes do without an account of the self. Korsgaard connected autonomy with the constitution of oneself; autonomous action is self-constituting action. We have discussed the problems with this proposal. O'Neill does not suffer these problems; acting badly (unautonomously) is misguided, and we should attribute blame for these bad actions practically. There is no self that 'ceases to be' when we act badly. We can view the person who acts badly, understand the action as wrong and meanwhile practically assign personhood to this creature. This is not made impossible by the bad action itself. We can conceive of the determinedly evil person on O'Neill's view. This would be exactly what we expect of such a being, someone who comes across as generally coherent and in control of his actions and yet capable of doing unspeakable evil.

Section 3.7: O'Neill's proposal; self-legislation

We are left with two somewhat diverging attitudes to dealing with 'Kantian' practical reason; a very substantial and a very lean account of the acting subject (and its freedom in acting). Both adhere to the demand of symmetry (the radically evil person is possible on both their accounts) and to the demand of asymmetry (both argue that the autonomous action is preferable to the unautonomous action).

[142] O'Neill, Towards Justice and Virtue, p.108

We believe however that there remains in O'Neill's account a gap, one that is filled by Kant due to his insistence on the importance of an account of the self. O'Neill's arguments make clear that it is sensible to *construct another's reason* by analysis of our own attitudes towards these others and these others actions. Here, the *fact of the matter* concerning people's metaphysical status as agents is indeed not required. But, can we understand *ourselves* in a similar *constructed* fashion?

For our own status, O'Neill's solution would demand of us to reconstruct who or what we are from what is implicit in our actions as well. But this is an awkward proposal, our actions can only be actions and have moral significance if they are undertaken self-consciously; we need to presuppose *in our own acting* already our status as a person. This is not a problem when dealing with another because it is perfectly acceptable to say that *from both my and his acting I conclude the other's personhood, even though were I to be correct, this means that his actions were already, before my construal, self-conscious and personal.* But, I cannot apply such a thought to myself, I cannot conclude my personhood from my acting after the fact and then state that *if I am correct, my actions were apparently already before the construal self-conscious and personal.* Because, if my action was self-conscious, this cannot be something that I was not aware of at the moment of acting itself.

Construal of personhood *after the fact* is impossible if the construer needs to construct his own personhood of a fact that needs to be self-conscious. And, if the action was not self-conscious, it is exactly to admit that it was not a proper action that can be stated to have moral significance. O'Neill's proposal to let pragmatic, or practical, considerations trump theoretical questions of personhood are well taken for so long as the status of others is of itself not accessible at all. But, when we consider our own status, that is (at least to the extent of being self-conscious in our acting or not) accessible, pragmatical reconstruction is no longer an option open to us.

How does this trouble O'Neill's account? It does so because it casts doubt exactly on O'Neill's insistence of the importance of autonomy of practical reasoning. It re-mystifies what Kant has attempted to clear up. In 1.5.2 and 3.5, we have seen Kant focus on the problem of *how* we can make sense of autonomy. Because, autonomy, or giving the law to oneself, is all but an easy to grasp concept. In 1.5.2 we saw that *at first sight, the concept of giving the law to oneself is contradictory*, for the one who obliges could at any time release the obliged from the obligation. But, an obligation cannot be merely up to us, it is not so that we are only obliged to be moral when we choose to understand ourselves as obliged to be moral. The concept of a duty would dissipate as a result. Because of this, Kant argues we *must* have an account of a person, split into law-giver and law-receiver. And in 3.5 we saw this further worked out, our

Willkür is given a law by the *Wille*, *our acting in the world should in the end adhere to the demands of reason* i.e. the law given by the *Wille* must be followed by the *Willkür*, and in effect generate morally acceptable actions in the world.

O'Neill ditches this construction. Autonomy consists in that the law for practical reason, derives from, and applies to, practical reason itself. This however, runs into exactly those problems Kant wanted to evade by splitting up law-following and law-giving into two parts of the same person. Now we can again wonder whether *if practical reason simply decides not to obligate itself, whether it remains unobligated*. It seems to be so, while Kant infuses *Wille* with the power of mere lawgiving (and necessitating the moral accurateness of this faculty, which will thus always oblige us), it is not giving the law to *reason*. It gives it to us, finite humans, to follow. O'Neill bites a bullet on this however, and is described in the Stanford article on *Constructivism in Metaethics* as telling us:

> *The process of figuring out what the principles of reason are is avowedly circular. This circularity is not worrisome because the process of verification is reflexive, as it involves reason critiquing the claims of reason itself.*[143]

O'Neill tells us that there is no contradiction, but mere circularity. Practical reason provides the law to itself, and indeed, should follow that law by its own decree. This requires however that practical reason *actually obliges itself*, and does not withhold this obligation from itself. *We need to accept ourselves as reasonable and principled creatures before we actually oblige ourselves to action in accordance with this constitution*. This saves autonomy, but makes duty awkwardly volitional. The *unworrisome circularity* that O'Neill asserts however can be avoided by a *more fleshed out account of autonomy*. O'Neill tries to argue the moral import of autonomous action without a recourse to the exact *working* of autonomy. The concept itself should give us pause, and we should conclude from it the conclusions we need for pragmatical, moral, purposes. How autonomy, how giving the law to oneself, works, remains relatively unexplored. The positives of this is that it does not need to go into the *metaphysical machinery* that makes autonomy possible. From a practical perspective, when considering duties to another, this is enough. But, the *avowed circularity* can be avoided by *unpacking* some account of the self that can make sense of

[143] Bagnoli, Constructivism in Metaethics

autonomous action and reasoning. This is the positive of Kant's account that provides us with the split self as a model for self-legislation.

Depending on what theoretical construct one finds more troubling, either metaphysical superfluousness or avowed circularity, one can prefer the deflationary or the substantialist view of the self we discussed. O'Neill's account of autonomy is, relative to Kant's account, a black box. The importance is that we understand *what is implied by autonomous action*, not *how autonomous action is even possible*. This is the story that Kant tells us. But, this is exactly because we have seen reasons for why autonomy needs to be unpacked, *at first glance it is either contradictory or circular*. This means that, although we might aim to give a less substantialist account of autonomy than Kant provides us with, still O'Neill's comfort in not discussing the way that *actual self-legislation* is possible is not a satisfying strategy.

Chapter 4: Extensions of the Self

How can we ourselves improve ourselves in an effective and just fashion?

Abstract: In this final chapter we will apply what we have come to understand of the Kantian model of the self. The aim is to investigate whether the model can be productive for finding methods of self-improvement. First we recap what we have come to know on the dual-self by comparing its implications for our attempts of attaining perfection in both the practical and the theoretical domain. After this, we begin with the most well-known method for self-improvement, ascetic exercise, and discuss Kant's views on this and the importance of caution when applying such exercise. Secondly, we will look at the importance of attaining knowledge of oneself for the project of self-betterment due to self-knowledge's transformative power. This finds its way into the importance not only of gaining a factually accurate understanding of the self, but also a productive aesthetic understanding of oneself. We discuss techniques of attaining self-knowledge and a productive aesthetic- narrative- or practical- understanding of oneself, both from modern and from Kantian perspectives. Finally, we will discuss Kant's views on the moral betterment of people through education and expand upon this by considering the ever increasing importance of dealing with oneself in a vacuum of *classical role-fulfillment*; bettering and designing oneself requires techniques that we should attempt to teach.

Section 4.1: Aiming for perfection

In the theoretical domain of reason, Kant argued that we are presented with questions that our reason in its theoretical use cannot hope to answer. Issues regarding the status of the soul, the (in)finitude of space and time and other metaphysical conundrums. Although unanswerable by reason, they are necessarily *posed* by reason.[144] We cannot remain indifferent to these questions. This was why Kant argued that we needed a critique and subsequently, as he argues in the doctrine of method, a *discipline* of reason.[145] Reason needs to refrain from trying to answer these questions, despite the impossibility of weeding out the wish for knowledge of this kind.

In the practical domain we are posed with the finitude of the human condition which problematizes the attainment of (moral) perfection. We may want to attain complete and certain mastery of ourselves, necessitating our moral behavior. This is the same thing as to say that the *principle of self-love* does not hold sway on our inclinations anymore, like was the case for a *holy will*. But this is then

[144] Kant, Critique of Pure Reason, BXIV-BXV/p.109
[145] Kant, Critique of Pure Reason, B739-B740/p.629

simply to deny that there are parts of ourselves that are distinct from our transcendental existence; it amounts to denying there are parts of myself that are *given* and not *spontaneous*.

We cannot simply decide to be angry, nor can we simply decide not to be inclined to do this or that, these are elements of our psyche that are given to us. We are always of course free to *decide our course of action*, but besides this, we are left adrift, prey to our own contingent constitution. Hellenistic writers were well aware of this struggle. In their attempt to attain *ataraxia,* or freedom of disturbance, we see how all the Hellenistic schools in their own way attempt to cast off the contingency of their own constitution. Analyzing what is *up to yourself* and what is not becomes an important task, as some propose to drop one's attachment to those things which are not up to oneself. Marcus Aurelius argues that pain can be overcome by simply *withholding judgement*[146] from the painful impression; we should only be our reason and disinherit our ownership of the rest of us. More in line with Kant's views we have Sextus Empiricus claim that *the painfulness of the impression (hunger, cold) is given and thus will be unpleasant*.[147] Yet, we are indeed free in our judgement of whether pain is unjust, so what we are capable of is escaping the *despair* that people in pain often experience.

We have discussed whether we can make sense of given impressions, habits, inclinations *being up to us* without claiming them to be *free and spontaneous* as our thinking and our acting is. We concluded that although we cannot will our impressions away, we can try to alter our constitution and through that detour, apply our will to that which is given. Like the body can be made more perfect by acting upon itself (eating well or applying *exercises* of various sorts), so can we act on our habits, inclinations and even the nature of our impressions. We cannot decide to be morally perfect, but we can work to inch closer and closer. Perfection is not something to be attained, it's something for which we should strive. The idea of perfection guides this striving because the actual *end* of our striving is to inch closer and closer to perfection.[148] And although we may never actually become morally perfect (we can never lay our empirical

---

[146] *"If you remove your judgement of anything that seems painful, you yourself stand quite immune to pain. 'What self?' Reason. 'But I am not just reason.' Granted. So let your reason cause itself no pain, and if some other part of you is in trouble, it can form its own judgement for itself."* (Marcus Aurelius, The Meditations, book 8; 40/p.75)

[147] *"We do not, however, take Sceptics to be undisturbed in every way – we say that they are disturbed by things which are forced upon them; for we agree that at times they shiver and are thirsty and have other feelings of this kind. But in these cases ordinary people are afflicted by two sets of circumstances: by the feelings themselves, and no less by believing that these circumstances are bad by nature. Sceptics, who shed the additional opinion that each of these things is bad in nature, come off more moderately even in these cases."* (Sextus Empiricus, Outlines of Scepticism, sec. xii/p.11)

[148] Kant, Critique of Practical Reason, 5:122/p.98-99

existence wholly beside us without ceasing to be), we can *strive* to become more perfect; *we owe it to ourselves*.

So, like theoretical reason leaves us wanting in our incessant wish for metaphysical knowledge, practical reason leaves us wanting in the image of perfection it presents to us but that we cannot attain. Both metaphysical knowledge and our perfection lay out of our reach, and they lie out of our reach precisely because we are empirical creatures who have a given character and share in the forms of sensibility. Yet there is a clear difference. The conclusions of the first *Critique* provide us with a *discipline* of reason, we need to attempt to refrain from attaining metaphysical knowledge. But the conclusion of Kant's practical philosophy is that we should substantialize the idea of '*striving for'* positively. Attainment of perfection is not normative, but striving for moving closer to perfection, is. This difference in attitude in the practical and theoretical domain is exemplified by Kant's turnabout regarding the antinomies of reason. Where the impossibility of knowledge of *the eternality of our soul* for example in the theoretical domain is understood *negatively* (there can be no knowledge of such things) in the practical domain we are allowed, *positively* to *postulate* precisely those objects; i.e. we may act *as if* they existed.

Over the course of the thesis, we have attempted a number of synthetizations between modern philosophical movements (virtue ethicists, neo-Kantians) and the classically Kantian picture. Sometimes, parts of these movement's annoyances and contributions meshed well with the Kantian structure we introduced in chapter one (understanding Kant from the viewpoint of his ideas on the self). We believe we showed that *virtue ethics* can readily be caught in a Kantian framework in a productive way. In the previous chapter, we argued that O'Neill's and Korsgaard's attempts to reinvigorate the Kantian moral program with a focus on, respectively, construction of personhood and self-constitution, forego a number of the advantages of orthodox Kantianism. Especially in the relation of the self to itself, we have argued that Kant's metaphysical machinery may be understood to be more than mere metaphysical superfluousness. But at the same time, through these synthetizations, our own understanding of the Kantian picture has grown beyond the standard reading as well, with a distinct focus on *a Kantian theory of self-improvement*. All the ways we have argued against or with, virtue ethicists, O'Neill and Korsgaard, has provided to us an understanding of relating to, and improving, the self that we believe not only to be correct, but also to be potentially productive. We invite anyone to read Kant's *doctrine of virtue* in the *Metaphysics of Morals*, which is a work filled with insightful examples and views on improving the self that provide handles for action (which is not to say Kant there never posits foolish advice and duties). We believe that the theoretical basis we have discussed throughout the thesis can however have many more

practical implications. We will provide applications of what we have discussed so far with the direct interest of providing insights that can be applied in one's relating to oneself.

Section 4.2: Ascetics

The first, very well known, activity to gain mastery over oneself and to improve oneself is through the broad ancient tradition of ascetical exercises. In section 2.5.2 we have quoted Kant's claim that one of the most important aspects of ascetic exercise is for it to be accompanied by a *cheerful heart*. Before we delve deeper into this we should see how Kant finishes this up, telling us about the counterpart of ascetic exercise accompanied by a cheerful heart, monkish asceticism:

> *On the other hand monkish ascetics, which from superstitious fear or hypocritical loathing of oneself goes to work with self-torture and mortification of the flesh, is not directed to virtue but rather to fantastically purging oneself of sin by imposing punishments on oneself. Instead of morally repenting sins (with a view to improving), it wants to do penance by punishments chosen and inflicted by oneself. But such punishment is a contradiction (because punishment must always be imposed by another); moreover, it cannot produce the cheerfulness which accompanies virtue, but rather brings with it secret hatred for virtue's command. – Ethical gymnastics, therefore, consists only in combating natural impulses sufficiently to be able to master them when a situation comes up in which they threaten morality; hence it makes one valiant and cheerful in the consciousness of one's restored freedom.*[149]

Kant's warning sounds, to a modern ear, perhaps somewhat dated. Of course *mortification of the flesh* and *self-torture* are not methods we would generally be drawn to. But, such techniques of the self are these days as alive as they were two hundred years ago. And although they are not as *visible* and *physically abusive* as they used to be, they continue to wreak the havoc Kant warns for.

---

[149] Kant, The Metaphysics of Morals, 6:485/p.245

The impulse to hold oneself accountable, or as sometimes stated '*having a conscious'*, and to beat oneself up over what one has done, are commonplace. And of course, feeling bad over what one has done is in some sense perfectly fine. One can *regret* what one has done. But, when this feeling of regret becomes too strong, becomes a proper *bad feeling*, and remains firmly entrenched in our psyche, we are tempted to understand this bad feeling as proper, just like we would for mere regret. The regret turned entrenched unpleasantness, can then be understood as punishment, and after the *allotted punishment has been lived through* we are freed of both the nasty feeling of regret and of the moral requirement to regret our action. But this oversells the idea of the nasty feeling as punishment. For, when issues of virtue are concerned, at no point should we consider punishment as a good means to an end; virtue is future oriented, and punishment can only be considered proper insofar as it is conducive to the generation of virtue. Furthermore, the punishment cannot undo the requirement to learn from the mistakes you make, to (intellectually) regret their passing, and to work to make sure it is not repeated. Virtue is not the same as justice or right. There, punishment can be dealt by external authorities, and after the punishment the outward transgression (*right*, or, *Rechtens,* concerns those duties we have towards others where we can be expected to adhere to those duties in a way that can be demanded by the others)[150] can be *forgiven*. But, virtue, although maybe aimed at others well-being, is focused on yourself; I must give the law to myself, and only I can judge for myself whether I did so. There is no forgiveness to be given, but this is because there should be no concept of punishment to begin with.

When one has procrastinated an entire day away, when one skipped on his diet or when one harbored feelings of jealousy, it is not only common, but even understandable, to want to apply punishment. For, after the punishment the *sin has been purged* and we can leave our previous actions behind. This punishment might consist in fasting, or becoming angry with oneself, or to harbor the unpleasantness of the knowledge of your un-virtuousness toward the brink of weeping.

But, not only can we leave our sin behind, we can tell ourselves that this is a truly virtuous action, for I punish myself in a way that will deter me from acting in such a way again. At this point we attempt to rationalize our self-inflicted punishment and then we properly find ourselves in the realm of monkish ascetics, as Kant calls it. And the problem is of course that this punishment cannot be conducive to virtue; for it breeds hatred of oneself and of that which you now associate with punishment, namely virtue.

[150] Kant, The Metaphysics of Morals, 6:229/p.26

We should not only be capable of being satisfied with ourselves, with our moral asceticism, but it should also be actually productive. And for *ethical gymnastics* to actually be effective, it needs to be accompanied by a mindset that is *productive in and of itself*. This is why Kant doubles down on the importance of cheerfulness. If our asceticism makes us more capable of withstanding hunger, but at the same time, makes us bitter and robs us of our enjoyment of food, not much has been gained, neither from a humane perspective nor from a moral one. Because one's bitterness will outweigh one's hardy constitution in moral conduct. This is not only the case for ascetics, but for every action that aims at developing oneself; it needs to be helpful both in the narrow aspect of oneself that is being developed, and in the broad whole of the self that is affected by the regimen of improvement as well.

Ascetic, and any other self-improvement focused, exercises should always be conducive to one's virtue. In this, one needs to seriously consider one's general mental status as well as the specific aspect of oneself that is being trained. There is a certain easiness to *focus on merely one aspect of yourself* when aiming to improve yourself. Partly, this is because it is complicated to take all the effects your actions have on yourself into account. However, partly, it is also because it is attractive to try and *compensate* for some of our flaws by improving and excelling in that which comes to us naturally. However, when we lose too much, leave our flaws to fester under the activities we undertake, our self-improvement cannot be deemed successful. This is particularly troubling for some forms of ascetic exercises, for asceticism that embitters is not worth the local strides one makes in self-improving. And asceticism that functions on the concept of self-inflicted punishment will embitter. Despite the attractiveness of *suffering for your faults* and then being done with it, it is not a virtuous way of acting. When one disappoints oneself one needs to both accept the disappointment, learn from it and yet, not become self-hating, harboring the bad feelings as the proper punishment.

We see the ugly side of these kinds of self-destructive strategies in Tolstoy's *War and Peace*, where Prince Andrej wants to test the strength of his love for the young Natashja, and her love for him. They decide to do this by separating for a whole year, to afterwards step into wedlock. But, this perverse test of course became a self-fulfilling prophecy; there was nothing lacking in their love that would have been troubling over the course of their potential wedlock. But the young Natashja was not up to the Herculean task of remaining faithful to *her perceived love for Andrej*, for of course, love wanes when saddled up with these sorts of tasks after only a short period together. What first seems like a prudent plan to test the depths of their love is little more than a ruse, for the test itself makes demands of the lovers that would not exist outside of the testing scenario. Afterwards, they both plunge themselves into

the depths of self-loathing, punishing themselves for their foolishness for the rest of the book, resulting in gloom that at no point finds its way into good action. Andrej becomes death-defying and Natashja reclusive. Both the ascetic practice to test beyond that which can be deemed useful (placing yourself in situations you might wish to be able to withstand, but which will not reasonably occur) and afterward applying punishment to oneself are bull-headed ways of acting that lead to destruction, not salvation.

The dangers of ascetic practice are both argued explicitly by Kant and are implicit in the story we just told on Andrej and Natashja. Self-improvement needs to be *actually effective*, and should not merely *play to the idea of improving oneself*. This means that concrete circumstances in the world ('am I not now preparing for a situation that will never come naturally by placing myself in it?') and your own psychological and physical state ('will this training make me sad, gloomy, sick or self-resenting and how can I change the exercise so that I do not have to be self-destructive in my training?') need to play a role in the consideration of ascetical exercise. Ascetic exercises are easily misused. For that reason it should be undertaken only when one has an overview of what one wants to accomplish with it. Ascetic exercises can have negative effects and doing them should be done with measure and in full awareness. We should aim to minimize self-flagellation, not only in its physical form but also in its mental form; through an understanding of its destructive, non-virtuous character. But, most importantly, any ascetic exercise needs to be prefigured by a good understanding one's own character, to minimize the odds of the exercises having effect on our constitution contrary to its goal.

Section 4.3: Transformative power of knowledge of the self

In the previous section, we pointed to the importance of having an accurate understanding of oneself to minimize the possibly determinantal effects of ascetic exercise. But, this undersells the importance of self-knowledge. For, to effectively act on oneself *at all*, we need knowledge of how to act. Our virtues, our bodily properties, our impressions, these are given to us; they resist the power of our will and are not up to us directly. However, they are up to us in a second-order form; I can act upon myself to change these given properties of myself. But, this means that to act on myself in such a fashion, I need more than the *end* of the action, I will also need to discover proper *means* (since willing the end is itself not sufficient for given properties). This means I need to know my *self* in its empirical form and its relation to the rest of

the world. I can only act effectively on myself when I know myself, so the project of coming to know oneself is a pre-requisite of the whole project of self-improvement. And not only in discovering the means is this knowledge necessary; for if I do not know who I am I cannot come to conclude I am not in some specific way unlike who I want to be. Whether or not some *abstract end* is actually already achieved is again a question of empirical knowledge of the self; I can only act so to rid me of my fear of heights if I already know (and acknowledge) myself to have such a fear, ignorance of this fact makes the end impossible to adopt.

In this direct sense, knowledge of oneself is *transformative*, it is the pre-requisite to changing oneself in accordance with what you want to become; for without knowledge of oneself, no self-improving ends will be salient nor will we be capable of ascertaining any means to achieve those ends. However, it is also transformative in a second way. For, we will have at all times some understanding of ourselves, and we will act on it whether our understanding of ourselves is completely accurate or not. This means in effect that, when our understanding of ourselves changes, we in effect, down the line, *transform* ourselves as well, as we act in accordance with that understanding. This is the topic of Michel Foucault's early period, the archeological period. In this period Foucault was interested in sciences that had the human as their object. He argued that these sciences produce knowledge of the human, or the subject, and so create a discourse that will be internalized by the humans (the objects of investigation) themselves. Subjects are constituted by the discourse that surrounds them, so science of humans changes a person's self-understanding and through that the person. He ends his major work of the archeological period, *The Order of Things: An Archeology of the Human Sciences*, with the following remarks:

> *One thing in any case is certain: man is neither the oldest nor the most constant problem that has been posed for human knowledge. Taking a relatively short chronological sample within a restricted geographical area – European culture since the sixteenth century – one can be certain that man is a recent invention within it. It is not around him and his secrets that knowledge prowled for so long in the darkness. In fact, among all the mutations that have affected the knowledge of things and their order, the knowledge of identities, differences, characters, equivalences, words – in short, in the midst of all the episodes of that profound history of the Same – only one, that which began a century and a half ago and is now perhaps drawing to a close, has made it possible for the figure of man to appear. And that appearance was not the liberation of an old anxiety, the transition into luminous consciousness of an age-old concern, the*

entry into objectivity of something that had long remained trapped within beliefs and philosophies: it was the effect of a change in the fundamental arrangements of knowledge.[151]

Man, in its specific instantiation that we find him right now, as (sometimes implicitly) described by the sciences of biology, economics and linguistics, came into being only recently. Our collective *understanding of man* and through that, *our collective understanding of ourselves*, changes not merely who we take ourselves to be, but also who we actually are.

We have identified two ways in which knowledge of oneself can be *transformative*. The primary transformative power derives from knowledge of oneself making it possible at all to act on oneself, let alone effectively. If our knowledge of ourselves is severely lacking, we might act on ourselves with an aim to produce an effect, while at the same time, in fact producing a contrary effect. A large pile of work on my desk is giving me an anxious feeling that I would like to shake off. I conclude that, since the pile is the cause of my anxiety, I attempt to work till deep in the night to reduce it. However, afterwards I sleep so poorly from over-exhaustion, that whatever strides I had made are lost due to ineffectiveness the next day, making my anxiety more firmly entrenched than it was. A more developed self-knowledge could have, in this concrete case, made it possible for me to actually (partially) attain my end, instead of working against my end from my own accord. But, we have seen Foucault point to a second transformative force of self-knowledge. This is a derivative transformative force that leeches on the first. The eventual transformation still takes place due to our actions being in accord with our knowledge of ourselves, but now we consider the factors that i) we always have some self-understanding and ii) that it is sometimes hard to point to a clearly *more accurate* self-understanding that is readily available. In these cases, the transformative force becomes *heteronomous*, we are caught in one or the other way of understanding ourselves without a clear way of improving or changing this self-understanding as it is deeply embedded in our more general ways of understanding the world.[152] This means that attainment of knowledge of oneself needs to happen on multiple levels. The first possibility is to look at the human sciences and apply their knowledge on your own person. We read psychology and biology and through that develop our understanding of our empirical selves. But, the derivative power of knowledge of the self shows also the

[151] Foucault, The Order of Things, p.421-422

[152] Or in Foucauldian terminology, the episteme we exist in defines our self-understanding in ways invisible to ourselves.

importance of attempting to attain knowledge of oneself in a less mediated fashion. We cannot expect all knowledge of ourselves to be easily accessible via empirical science due to the self-fulfilling character of broad understanding of the person already implicit in their way of thinking.

Foucault believed so too, and in his final period, he argued for the possibility of taking our self-understanding into our own hands. For Foucault this took the shape of finding new (or forgotten) ways of attaining self-knowledge that are less dependent on structures outside of us. Foucault's focus is on regaining *techniques of the self* that have over time been constructed and then again have been forgotten. This means that a lot of Foucault's work in this period is historical, despite the project's explicitly philosophical nature. Ancient *spiritual exercises* are the bread and butter; analyzing the techniques of Stoics, Cynics, Epicureans, Plato and Aristotle for getting to know oneself properly. Let us discuss one example of these techniques that Foucault proposes to help in the acquisition of self-knowledge in a way where the knowledge is properly dependent on the true nature and the activity of the inquisitive self. There are of course many more possible techniques for doing this and we do not claim to be exhaustive of Foucault's proposals nor all the possible techniques there might be.

In *The Courage of the Truth* Foucault interprets Socrates' specific way of *speaking the truth* (*parrhesia)*, a candid and courageous way of speaking, to have as aim to *help others help themselves*. At first, *parrhesia* is discussed in the context of politics. In politics, *parrhesia* points to the courage required to speak to the assembly of the city in a way where one is not afraid of the fact that the people*,* will be unhappy with the *ugly truth* that one needs to speak. This courage for truth is required for the prolonged existence of a city-state because it is only through it that it is possible for the city state to get to know itself properly. If every investigator of from the city itself is afraid to candidly speak the truth about the city, the city will never gain any self-knowledge.[153] In the end, the conclusion is that *parrhesia* in a political context is so unstable, demands so much of the speaker and is so ill received by the listener we should do away with attempting to attain it. In the personal case however, it is far more promising.

For the person, something similar is the case as for the city state. We lack the courage to believe the truth that is presented to us (we are the people who would rather not hear the painful truth about our constitution) and we lack the courage to speak the truth to ourselves forcefully (we are not the

---

[153] Foucault, The Courage of Truth, p.40-41

courageous speaker, and simply pander to the people inside of us). Foucault states that what Socrates is aiming to help us with is:

> *This other aim is in fact to see to it that people take care of themselves, that each individual attends to himself as a rational being having a relation to truth founded on the very being of his soul. And in this we now have a parrhesia on the axis of ethics. What is at stake in this new form of parrhesia is the foundation of ethos as the principle on the basis of which conduct can be defined as rational conduct in accordance with the very being of the soul.*[154]

There are three important characteristics of the ethical form of *parrhesia.* First, everybody should *see to themselves*; instead of raising others (a speaker of the city-state to its population), we should raise ourselves. Second, this care for yourself should be *founded on the way your soul is*. Third, this should find its way into conduct that can be deemed rational *given the nature of your soul*. The aim of *parrhesia* is then, in our terms, to develop an accurate understanding of yourself, to do so yourself, and to put this into action. The reason we need Socrates to teach us to apply *parrhesia* to ourselves is because this courage to *come to terms with the very being of your soul* is not always pleasant. We can both be fearful of speaking candidly against ourselves and we might be unwilling to listen to ourselves. Socrates attempts to lead by example and intellectual insight.

However, there are luckily more concrete ways in which we can compensate for our divergence form the *ethical ideal figure of Socrates*, and ways for us to cope with our innate *vice* of *cowardice*. In *The Hermeneutics of the Subject,* Foucault traces the Hellenistic practice of writing letters to a trusted friend or to oneself.[155] The letters are written for the purpose of examination; the writer expounds the status of his mental and physical life on the letter, and the recipient is supposed to examine the person described on the virtue/vicehood of his character. The practice allows to separate the three voices that one should be at the same time. Fully autonomous, we should both speak, evaluate and accept the truth about ourselves to ourselves. This is tough, both intellectually, one needs to differentiate between three voices in one instant, and because we can dislike the message. And when we dislike the message, it is easy to

---

[154] Foucault, The Courage of the Truth, p.86

[155] Foucault, Hermeneutics of the Subject, p.89, 162-164

discard the need to listen because of the obfuscation provided by the awkward one person-three functions setup. However, when we write a letter, we are providing the reader with knowledge of our person, without the need for ourselves to evaluate this, nor the need to accept its complete accuracy. In the case of the other reader, we are presented with an evaluation of our person that was based on our own account of ourselves. It is no longer an option to retract the earlier account, and we are now more acutely pressed to accept the truth that is laid before us. If the letter was send to ourselves, we still need to do the evaluation and the acceptation of the truth in one and the same breath. However, again the account itself that has been set out before us can no longer be retracted, making it more pressing to accept the account and take it serious in our understanding of ourselves. In these cases we are gaining knowledge of ourselves *more or less*, by ourselves. Many more techniques for attaining self-knowledge are abound, and getting them clearly in view will aid anyone in their broader project of self-improvement. For if self-knowledge is the pre-requisite for any self-improvement, and accurate self-knowledge is pre-requisite for any effective self-improvement, we should keep our eyes peeled for any and all practices that can aid us in getting to know ourselves.

Section 4.4: Narrative understanding of the self and practical identity

In the previous section we discussed the transformative force that inheres in knowing oneself to be one way or another. We mostly considered questions of *factual knowledge of oneself* and how one can take the acquisition of that knowledge into his own hands. However, there is a less *factual* way in which we can make sense of having *self-understanding*. This is when we *understand ourselves to adhere to one or another archetype*. Or, when we have as Korsgaard calls it, a practical identity:

> *Conceptions of practical identity include such things as roles and relationships, citizenship, memberships in ethnic or religious groups, causes, vocations, professions, and offices. It may be important to you that you are a human being, a woman, or a man, a member of a certain profession, someone's lover or friend, a citizen or an officer of the court, a feminist or an environmentalist, or whatever. Our conceptions of our practical identity govern our choice of actions, for to value yourself in a certain role or under a certain description is at the same time to*

find it worthwhile to do certain acts for the sake of certain ends, and impossible, even unthinkable, to do others.[156]

This kind of self-understanding is different from what we discussed in the previous section. It is not merely descriptive, what we believe to be true of us which in effect changes how we act on ourselves, but also (quasi-)normative. It influences all of our actions because *due to our understanding ourselves as being this or that archetype we come to understand certain actions as proper to us*. We can see this play a role in many facets of our life. When we buy clothes for ourselves we consider not merely *the beauty of the clothes themselves* nor do we only consider *the use and comfort of the clothes*; more often than not we also choose clothes that we *believe to be expressive of who we take ourselves to be* or in other words *we often choose clothes that are proper to who we are.* Another example from my own memory is when I as a young child, from seeing *cool* children on TV disliking spinach[157] (and understanding myself to be a similar child), told my parents (who presented me with spinach on my plate) that I do not like the taste of spinach. My parents responded that I could not know this, for I had never eaten spinach. I responded that *I am not the sort of child that likes spinach*, disregarding the evidence that I in fact might be. Of course, after such a direct confrontation I might, and eventually was, pushed, to try the spinach, to find out whether I really was who I believed myself to be. There is in these cases and awkward conjunction between the descriptive (this is who I am) and the normative (this is what is proper for me).

As Korsgaard sees it, having these *contingent* practical identities is necessary to come to any action at all. As we discussed before in 1.5.2, morality severely underdetermines action, this was why we argued that we could not make sense of a holy will that acted. Korsgaard has a similar concern with the person that lacks any practical identity, for then that person would never find any action *proper for her to execute*. Korsgaard goes so far as to claim that this can figure in an argument for one or another practical identity:

Suppose now that I conform to my obligations as an American citizen, treating the duties of citizenship as duties to which I indeed must conform. Someone might say to me: okay, sure, I see

[156] Korsgaard, Self-Constitution, p.20

[157] And this must be an implicitly understood phenomenon. The cartoon Popeye had as its specific goal exactly to create the opposite effect, to make spinach appear as *an awesome vegetable*.

that you must do that insofar as you identify yourself as an American citizen, but why must you take that way of identifying yourself so seriously? It's only an accident that you were born in America. And here part of the answer is that I must take some ways of identifying myself seriously, or I won't have any reasons at all. Insofar as I take that fact – the fact that I need some way of identifying myself – to be a reason, I express the value I set upon myself as a human and rational being.[158]

Although we agree with the claim that we need some contingent constitution, we disagree with the claim here by Korsgaard that this knowledge can figure in our defense of our practical identity that has been brought up by some interlocuter. This is to assume a certain *conservatism* regarding our practical identity; we would claim that not only can we set our practical identity aside (as Korsgaard allows for as well) we are also capable of *breeding* a practical identity. When we are confronted with the contingency of our practical identity the only way Korsgaard's argument could help us is if we are incapable of changing our practical identity. But, if we allow for the possibility of changing one's practical identity, the fact that we need one only tells us that if we accept the problematic (or just lackluster) character of our current practical identity, we should aim to conceptualize and eventually attain, a new one. The wildly interesting phenomenon of practical identity is not merely a tool for *understanding and describing* the way we come to prefer certain ends over others. It is also a tool to become the sort of person who we want to become; in this case, practical identity turns into *narrative exercise*.

For this, step one is to make explicit our practical identity. Where our mere contingent *reasons for actions stop* and our *contingent practical identity turning into reason for actions, begin*. Our implicit practical identity might be made explicit and self-conscious at least in the way that Heidegger believed all or most awareness functioned; when being confronted with their inability and *brokenness*.[159] In these cases it is the world that resists your self-image, either because it is abhorrent to the world itself, or because it is very clearly untrue and leads to inconsistencies in action. In such cases, you are also sabotaged in becoming who you understood yourself to be, for you can no longer justify your actions that would eventually make you so through habituation by recourse to your self-understanding. This will

[158] Korsgaard, Self-Constitution, p.24
[159] Heidegger, Being and Time, p.68-69

prompt you to reinvent yourself ad-hoc. This is what happened to myself when I *identified as a child who was cool and did not like spinach*; the world resisted my practical identity, requiring me to reinvent myself.

But this is not the only possible attitude. Through reflection we can also attempt to understand what drives us and who we understand ourselves to be. In those cases, we can also move beyond mere ad-hoc reinvention. We can try to take our self-understanding or practical identity into our own hands. The dual structure of the self allows for this; we can reflect from afar upon that object which we are. When we do this, we can attempt to change our self-understanding in such a way that it more accords with the person who we *want to become*.

In doing so, not all options are open to us. When we self-consciously are considering who we are, we are bound by laws of intellectual integrity to not attempt to become something by telling ourselves that we are something that we know ourselves not to be. This is what is so awkward in Pascal's proposal to become religious by acting as if we were; for in this case we know ourselves not to be what we try to convince ourselves we are.[160] The only option that is then left open is to tone back the self-aware discourse and *just do it*.[161]

What is left open for us however, is to understand ourselves differently *aesthetically*. Often, aesthetic categories and archetypes are understood vaguely, and multiple interpretations of the same facts are open to us, a number of them reasonable. This means that when we take our own lives and self, and apply some specific interpretation to that which we find, we might be free to understand ourselves in a number of aesthetic ways, all of which are reasonable. In that case, we can take into consideration which self-understanding would bring us closer to the (moral) ideal we have for ourselves and our lives. Narrativity seems particularly apt for this use, as it both allows for reinterpretation and yet is also prescriptive for the activities of the persons that occur in the narrative. Writers sometimes say that from some point onward, a book writes itself. The theme has been set, the destinies for the characters already written. Now it is merely a question of going through the motions, letting the characters act *as they should* to finish the story up. There is a quasi-normative dimension to any narrative, given a certain setup, things

---

[160] Pascal, Thoughts, p.39-40

[161] We suggest that Kant might have Pascal in mind specifically when he writes: *"Someone tells an inner lie for example if he professes belief in a future judge of the world, although he really finds no such belief within himself but persuades himself that it could do no harm and might even be useful to profess in his thoughts to one who scrutinizes hearts a belief in such a judge, in order to win his win his favor in case he should exist."* (Metaphysics of Morals, 6:430/p.197)

should come to exhibit what they in the end actually came to exhibit. This can be action-guiding when one narrativizes one's own activities. As the scholar of myth Joseph Campbell tells us about the function of myth in his seminal *The Hero with a Thousand Faces:*

> *The tribal ceremonies of birth, initiation, marriage, burial, installation, and so forth, serve to translate the individual's life-crises and life-deeds into classic, impersonal forms.* ***They disclose him to himself, not as this personality or that, but as the warrior, the bride, the widow, the priest, the chieftain****; at the same time rehearsing for the rest of the community the old lesson of the archetypal stages. All participate in the ceremonial according to rank and function. The whole society becomes visible to itself as an imperishable living unit. Generations of individuals pass, like anonymous cells from a living body; but the sustaining, timeless form remains. By an enlargement of vision to embrace this superindividual, each discovers himself enhanced, enriched, supported and magnified. His role, however unimpressive, is seen to be intrinsic to the beautiful festival-image of man – the image, potential yet necessarily inhibited within himself. […]* ***the man or woman who can honestly say that he or she has lived the role – whether that of priest, harlot, queen, or slave – is something in the fullest sense of the verb to be****.*[162]

Of course, fulfilling a role like the ones Campbell describes here was often done in a way heavily regulated by the surrounding community. But in our current day and age, we are less and less inclined to accept these *heteronomous* determinations of ourselves and our self-image. This leaves open the space to take it upon yourself to create your own narrative role; of course still in accordance with when, where and with whom you are, but freely within the limits of reality and intellectual integrity. We can construct our narrative identity and over time, construct ourselves through that.

Narrative exercise thus would consist of attempting to cast off the modern pessimism where we understand our own lives as merely chance driven chaos, or to cast off the traditionalism of pre-given narrative roles to fulfill. For Korsgaard, the contingency of our practical identity is a completely unproblematic feature of it. We would like to claim that, like virtue not being up to us in the direct sense and through that, we are not responsible directly for who we are, so also what practical identity we find

[162] Campbell, The Hero with a Thousand Faces, p.331

ourselves with is not our responsibility. However, what is our responsibility is to aim to improve ourselves and in effect, take our practical identity into our own hands. We become ourselves the creators of our life-story; not merely after the fact, *re*-constructing what we have done into a grand narrative, but during our actions and considerations, where we let our actions be guided by the *narrative we are constructing right now, that narrative that allows us to grow into the person we want to eventually become*.

Section 4.5: Moral Education

Despite Kant's reputation as a cold and loveless man (which we take from Manfred Kuehn's biography *Kant* to be partially true and partially false), his love and admiration for the education of children shines through. Most telling of this were Kant's efforts to support a new, progressive school:

> *Kant was always interested in education. This was not just because of his reading of Rousseau's Emile during the early sixties; it was something that he had worried about at least from his time as a Hofmeister. His lecture given at the becoming of Magister was, after all, "Of the Easier and More Thorough Presentation of Philosophy." In 1774, this interest in education received a new impulse through Johann Bernhard Basedow's (1723-90) founding of a new progressive school, the Philanthropinum in Dessau. The Philanthropinum was conceived in a "very progressive" spirit.*[163]

He went out of his way to support the school wherever possible. He donated money and collected money among colleagues for the school. He wrote articles aimed at recommending the school, and send promising student's to the school to take on teaching duties.[164] Kant subscribed very much to the Enlightenment ideals that the school was implementing; reducing the importance of religious education,

[163] Kuehn, Kant; A Biography, p.227
[164] Kuehn, Kant; A Biography, p.229

incorporating play into learning activities, and in general aiming to educate the students to "become independent citizens who could *take care of themselves* in their future lives."[165]

Kant has written on pedagogy specifically but for our current purposes we will keep to the remark that Kant has on moral education in the *Metaphysics of Morals*, because here his moral philosophy and general philosophical outlook feed very explicitly into his views on education:

> *The very concept of virtue already implies that virtue must be acquired (that it is not innate); one need not appeal to anthropological knowledge based on experience to see this. For a human being's moral capacity would not be virtue were it not produced by the strength of his resolution in conflict with powerful opposing inclinations. […] But since one does not acquire the power to put the rules of virtue into practice merely by being taught how one ought to behave in order to conform with the concept of virtue, the Stoics meant only that virtue cannot be taught merely by concepts of duty or by exhortations […]*[166]

Teaching morality, and teaching virtue, takes a strange in between position. For on the one hand, we all have our transcendental selves, that allows us to ascertain the moral law on the basis of the mere application of our pure reason. This means that the moral necessity of acquiring virtue is something that does not need to be told to us to do. What's more, since the normative force derives from our *autonomy*, that is to say, our giving the moral law to ourselves, it is out of the question to be handed correct moral conduct by others as authoritative. Yet, as we see Kant claiming, virtue does need to be acquired. So, teaching morality gets into an awkward situation where exhortations are no option. This leads Kant to the following proposal for teaching virtue:

> *For the* ***beginning*** *pupil the first and most essential instrument for teaching the doctrine of virtue is the moral catechism. […] A pure moral catechism, as the basic teaching of duties of virtue […] can be developed from ordinary human reason and (as far as its content are concerned) it need only be adapted to rules of teaching suited for the earliest instruction. The formal principle of*

[165] Kuehn, Kant; A Biography, p.227
[166] Kant, Metaphysics of Morals, 6:477/p.239

> *such instruction does not, however, permit Socratic dialogue as the way of teaching for this purpose, since the pupil has no idea which questions to ask; and so the teacher alone does the questioning. […] So the way of teaching by catechism differs from both the dogmatic way (in which only the teacher speaks) and the way of dialogue (in which both teacher and pupil question and answer each other).*[167]

The traditional catechism was introduced to expound religious doctrine to students; it took the form of *a question to be posed*, followed by an *answer that had to be memorized*. Both the question and the answer are given by the teacher, or by the manual from which the child learns. Kant adjusts this teaching aid to a form proper to moral issues. Because the tenets of morality can always be ascertained by the application of mere reason, we can expect the student to provide the answers to the questions themselves. However, virtue must, and can be, taught; this is done through the application of the right questions (which excludes true dialogue, there are a number of questions the student is supposed to mull over, and this should be the extent of the discussion).[168]

However, we would not be impressed by the example Kant follows to give of how a moral catechism might proceed. The questions are so leading, and moreover, the difficulty of the terminology and phraseology would make any modern reader respond that this is hardly proper didactics. Nothing of what the child would be forced to state would have been understood if done in this fashion.

Despite this, Kant's outlook is useful. Moral education should on the one hand always assume, upfront, the capacity of the one to be educated to come to understand the moral law by mere application of his reason. To assume otherwise would be to assume that the moral law would hold no sway on the student. To teach otherwise, would imprint the moral law into the learner as something that is to be taken from the outside, as a heteronomous law, while the law derives its lawfulness form its autonomy. Yet, we should teach morality in the way of *reaching out to others who are capable of understanding but must be led out of the fog in which they make clouded judgements.*

---

[167] Kant, Metaphysics of Morals, 6:478-479/p.240

[168] The famous scene of the slave who is taught mathematics in Plato's *Meno* comes to mind here. There also we find that the teaching should be made so as to be proper to specifically *remembering* what one had experienced in the realm of forms. This means that here as well, instruction and expositions were not proper modes of teaching.

But this reaches beyond merely the moral catechism. We can indeed help others ascertain the moral law by posing the questions to his reason that would bring into clearest daylight the commands of reason itself. We can also however, aim to help the student in more concrete ways. In the three previous sections we have discussed about possible exercises for attaining virtue and self-mastery. Narrative, ascetic and self-knowledge exercises that help you to capably understand what kind of subject you are and whether that is the sort of subject you want to be. All three of these activities are intellectually thorny issues; although we might expect reason to ascertain the moral law, we are foolish to expect everyone to be capable of ascertaining the details of effective self-mastery. Again, we can expect them to understand their importance, but not the details of their execution.

So besides 'teaching the good' through suggestive questioning, we can also properly *teach ways to become good* through instruction in the topics of narrative identity, asceticism and generating self-knowledge. Again we must be careful, ascetic practic*ing* and narrative exercis*ing* need in the end again be applied by the person herself again. This is what makes them interesting, they need to be executed by the person herself, so that it can be understood as the striving for (moral) perfection that reason tasks us with. To do otherwise would be to simply *discipline* the students into one or another proper form of subjecthood. We should teach the students to become self-aware, and knowledgeable of themselves to an extent where they become capable of applying self-techniques that allow for further virtuousness.

This could concretely take the form of asking primary or secondary school children to *summarize who they are* and then proceeding to discuss *why they take themselves to be so and so* and *comparing this to the views others had of you*. In higher classes one could *discuss one on one what kind of properties of oneself people are discontented with* and to *wonder what kind of action might alleviate the perceived problem*. In some cases this might take the form of *ascetic practice*; the person who wants to not go to sleep to late on Sundays arriving at school tired, might formulate the ascetic plan to not go to sleep late on Friday and Saturday anymore. In such cases, it would be important to point to the dangers of denying yourself transgressive pleasure at every turn of life's way. For narrative exercise first and foremost the reading and class-wide discussion of literature is imperative. In primary school one of the more productive questions one could ask for narrative exercise is the *why* question. Why did the things go as they went, why did person A do this and not that? Why did the story end so happily? Through this, even primary school students might be given a taste for a meta-understanding of literature, which will in time allow them to narrativize their own activity. In secondary school this might eventually lead to more personalized exercise. Who of these characters resembles you or you classmates? What was commendable or

reprehensible in this or that character? Story writing can aid as well. In the end the question *who are you* can be put to the student in an explicitly narrative fashion *what is your story, can you tell me multiple stories about yourself that show you to be different?*

All these self- and story-reflective exercises are not direct moral education. As we saw, direct moral education is necessarily very lean when one takes a Kantian approach to morality. However, providing tools that aid in moral self-construction, i.e. indirect moral education, opens up many concrete and varied ways of aiding the beginning student. And in the modern day and age, with the crumbling of traditional *life-fulfilling* roles as Campbell characterized them, not only Kantians, but philosophers of many moral persuasions, should positively react to the project of self-construction. In between *post-modern pessimism* which leaves us at best with cynical fatalists and at worst with radical inactivity, and reactionary traditionalism, which leaves us at best with small, tight, anti-cosmopolitan communities and at worst with large totalitarian states, it becomes difficult to uphold the ideals of the Enlightenment in one's own character. Amidst this tug-of-war, we would do well to help people as much as possible in their project of moral self-construction and cheerful self-fulfillment.

Conclusion

In the middle of Melville's *Moby Dick* we find a dismissive reference to the philosophers Locke and Kant. The second mate Stubb and the third mate Flask are returning with a *Right whale* which is not useful for the whale hunters' primary objective (it cannot provide them with the whale oil they seek), but will be used as a counterweight to the *Sperm whale* they already have hanging on one side of the ship. Melville/Ishmael tells us about an analogous activity that many philosophers (and scholars in general) might be well acquainted with:

> *In good time, Flask's saying proved true. As before, the Pequod steeply leaned over towards the sperm whale's head, now, by the counterpoise of both heads, she regained her even keel; though sorely strained, you may well believe. So, when on one side you hoist in Locke's head, you go over that way; but now, on the other side, hoist in Kant's and you come back again; but in very poor plight. Thus, some minds for ever keep trimming boat. Oh, ye foolish! Throw all these thunder-heads overboard, and then you will float light and right.*[169]

One can enter into deep, extended deliberation of a philosopher's system, only to eventually find some faults or a one-sided focus. In these cases, it is tempting to incorporate the thoughts of another philosopher's system into the previous system, balancing out the one-sided focus of the original system. We have done a similar thing throughout this thesis. We started by expounding the Kantian system, from the perspective of Kant's philosophizing about the self. However, in the stereotypical understanding of Kant, there is a one-sided focus on the primacy of reason and the good will in his characterization of the self. We believe this understanding to be mistaken, and have, by synthesizing insights from virtue ethics and neo-Kantians, attempted to enrich the Kantian framework and vice versa. We have counterbalanced Kant's view on the self by investigating what is of value in virtue ethics, and at the same time, have counterbalanced one-sidedness in virtue ethics with the Kantian framework.

[169] Melville, Moby Dick, p.272-273

But what Melville tells us is that although this might work to straighten out the problems one has encountered in one or the other thinker's system, we become more and more burdened with knowledge and we are left with something of a patchwork of philosophies. This patchwork is large and cumbersome and when managed poorly, renders a thinker inflexible, caught up in a web of intricacies with no clear direction or *theme*. The added benefit of learning from another philosopher through deep engagement with his or her system can be undone by a loss of overview and undue willingness to continue to defend that which one has accustomed oneself with. This is according to Melville why we should 'throw the thunderheads overboard'; it is the way to be both balanced and flexible, like the ship when unburdened by any whale carcasses.

I have myself felt burdened by these remarks of Melville throughout the writing of the thesis; how could I make sense of what I was doing as more than a 'balancing act', to get to a stable sort of Kantianism? The answer lies in that I believe more happens than a mere *balancing out*, the comparisons and synthetizations lead to insights and proposals that were not already implicit in any of the prime ingredients. In chapter four, we have expounded some of these insights and proposals that were not (all) already clearly present in any of the discussed literature, but can be understood to follow from thinking through- and comparison of- that literature. If that is so, philosophy's insistence on deep knowledge of philosophers of bygone days is not as misplaced as Melville takes it to be.

The Kantian system, with its insistence on a self that must be understood as *split* between the empirical world and the supra-empirical, noumenal, world, seems at first to many modern philosophers as an unpleasantly substantialist and metaphysical account of the self. If anything, we hope to have conveyed the *merits* to such an understanding of the self. We have made explicit its function in Kant's own philosophy (the endeavors of the first *Critique* and his moral philosophy would fall flat without recourse to an idea of *unalterable* subjecthood ingrained in us all). But we have also shown how much more modern concerns: of virtue, of self-care, of proper motivations, of moral-education, can be given an insightful new dimension when understood from the perspective of this *split self*. And in some ways, the structure of this split self has been my own attitude to a number of the problems we have tackled; for it is misguided to construe these problems as *naturalistic*, or *anti-metaphysical* areas of investigation (we must always, also when dealing with moral-education and virtue, keep the eternal in view). Yet, the fully denaturalized explanation would be misguided all the same, as it excludes parts and considerations concerning these problems that simply are not related in any way to the eternal and are given and contingent. To deny these would be just as one-sided. In this sense, we felt exactly like the Pequod,

balancing metaphysical plenitude and naturalist explanation. However, we did not feel burdened and heavy like the Pequod; what followed was for us not a patchwork system, but a new and fresh look on these already ancient problems.

Bibliography

Allison, H. (1990). *Kant's theory of freedom.* Cambridge: Cambridge University Press.

Allison, H. (2006). Kant on Freedom of the Will. In P. Guyer, *The Cambridge Companion to Kant and Modern Philosophy* (pp. 342-380). Cambridge: Cambridge University Press.

Anscombe, G. (1958). Modern Moral Philosophy. *Philosophy, 33*(124), 1-16.

Anselm. (1998). Proslogion. In B. Davies, & G. R. Evans, *Anslem of Canterburry; The Major Works* (pp. 82-104). Oxford: Oxford University Press.

Arkani-Hamed, N. (2012). The Future of Fundamental Physics. *Daedalus*, 53-66.

Aurelius, M. (2006). *Meditations.* Penguin Classics.

Bagnoli, C. (2017, Winter). *Constructivism in Metaethics.* Retrieved from Stanford Ecyclopedia of Philosophy: https://plato.stanford.edu/archives/win2017/entries/constructivism-metaethics/

Barnes, J. (1982). *The Presocratic Philosophers.* New York: Routledge.

Campbell, J. (2008). *The Hero with a Thousand Faces.* California: Joseph Campbell Foundation.

Copernicus, N. (1976). *On the Revolutions of the Heavenly Spheres.* (A. M. Duncan, Trans.) Devon: David and Charles (Holdings) Ltd.

Denis, L. (2006). Kant's Conception of Virtue. In P. Guyer, *The Cambridge Companion to Kant and Modern Philosophy* (pp. 505-537). Cambridge: Cambridge University Press.

Ferrarin, A. (2015). *The Powers of Pure Reason: Kant and the Idea of Cosmic Philosophy.* Chicago: The University of Chicago Press.

Fichte, J. G. (1982). *The Science of Knowledge.* (P. Heath, & J. Lachs, Trans.) Cambridge: Cambridge University Press.

Foucault, M. (1970). *The Order of Things.* London: Routledge.

Foucault, M. (2005). *Hermeneutics of the Subject: Lectures at the College de France 1981-1982.* (G. Burchell, Trans.) New York: Palgrave MacMillian.

Foucault, M. (2011). *The Courage of the Truth: Lectures at College de France 1983-1984.* (G. Burchell, Trans.) London: Macmillian Palgrave.

Heidegger, M. (1996). *Being and Time.* (J. Stambaugh, Trans.) New York: State University of New York.

Heraclitus. (2001). Fragments. In J. Barnes, *Early Greek Philosophy* (pp. 48-73). London: Penguin Books.

Hume, D. (2004). *An Enquiry Concerning Human Understanding.* Mineola: Dover Publications, Inc.

Hursthouse, R. (1999). *On Virtue Ethics.* Oxford: Oxford University Press.

Hursthouse, R., & Pettigrove, G. (2016, Winter). *Virtue Ethics.* Retrieved from Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/win2016/entries/ethics-virtue/>

Kant, I. (1894 reprint 2014). *Dissertation on the form and principles of the sensible and the intelligible world.* (W. Eckhoff, Trans.) Ediciones EK.

Kant, I. (1998). *Critique of Pure Reason.* (P. Guyer, & A. Wood, Trans.) Cambridge: Cambridge University Press.

Kant, I. (2002). *Groundwork for the Metaphysics of Morals.* (A. Wood, Trans.) London: Yale university Press.

Kant, I. (2004). *Prolegomena to Any Future Metaphysics* (Revised ed.). (G. Hatfield, Trans.) Cambridge: Cambridge University Press.

Kant, I. (2006). *Anthropology from a Pragmatic Point of View.* (R. Lauden, Ed., & R. Lauden, Trans.) Cambridge: Cambridge University Press.

Kant, I. (2010). *De religie binnen de grenzen van de rede.* (G. v. Eekert, W. v. Herck, & W. Lemmens, Trans.) Amsterdam: Uitgeverij Boom.

Kant, I. (2015). *Critique of Practical Reason* (Revised ed.). (M. Gregor, Trans.) Cambridge: Cambridge University Press.

Kant, I. (2017). *The Metaphysics of Morals.* (M. Gregor, Trans.) New York: Cambridge University Press.

Kirk, G. S. (1951). Natural Change in Heraclitus. *Mind, 60*(237), 35-42.

Korsgaard, C. (1996). *The Sources of Normativity.* Cambridge: Cambridge University Press.

Korsgaard, C. (2008). From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action. In C. Korsgaard, *The Constitution of Agency* (pp. 174-206). Oxford: Oxford University Press.

Korsgaard, C. (2008). *The Constitution of Agency.* Oxford: Oxford University Press.

Korsgaard, C. (2009). *Self-Constitution.* Oxford: Oxford University Press.

Kuehn, M. (2001). *Kant: A Biography.* Cambridge: Cambridge University Press.

Longuenesse, B. (1998). *Kant and the Capacity to Judge; Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason.* (C. T. Wolfe, Trans.) Princeton: Princeton University Press.

Longuenesse, B. (2017). *I, Me, Mine; Back to Kant, and Back Again.* Oxford: Oxfrod University Press.

Marx, K. (1969). Theses on Feuerbach. In K. Marx, & F. Engels, *Karl Marx and Frederick Engels Selected Works* (pp. 13-15). Moscow: Progres Publishers.

Melville, H. (1993). *Moby Dick.* Hertfordshire: Wordsworth Editions Limited.

Nagel, T. (1997). *The Last Word.* New York: Oxford University Press.

O'Neill, O. (1989). *Constructions of Reason: Exploration of Kant's Practical Philosophy.* Cambridge: Cambridge University Press.

O'Neill, O. (1992). Vindicating Reason. In P. Guyer, *The Cambridge Companion to Kant* (pp. 280-308). New York: Cambridge University Press.

O'Neill, O. (1996). *Towards Justice and Virtue: A constructive account of practical reasoning.* Cambridge: Cambridge University Press.

O'Neill, O. (2004). Kant; Rationality as Practical Reason. In A. Mele, & P. Rawling, *The Oxford Handbook of Rationality* (pp. 93-109). Oxford: Oxford University Press.

O'Neill, O. (2013). *Acting on Principle; An Essay on Kantian Ethics* (Second ed.). Cambridge: Cambridge University Press.

Pascal, B. (2002). *Thoughts.* Grand Rapids : Christian Classics Ethereal Library.

Pinkard, T. (2002). *German Philosophy 1760-1860; The Legacy of Idealism.* Cambridge: Cambridge University Press.

Plato. (1992). *Theaetetus.* (M. J. Levett., Trans.) Indianapolis: Hackett Publishing Company, Inc.

Plato. (1997). Cratylus. In J. Cooper, *Plato: Complete Works* (C. D. Reeve, Trans., pp. 101-157). Cambridge: Hackett Publishing Company, Inc.

Plato. (1997). Meno. In J. Cooper, *Plato: Complete Works* (G. Grube, Trans., pp. 870-897). Indianapolis: Hackett Publishing Company, Inc.

Plato. (1997). Timaeus. In J. Cooper, & D. Hutchinson, *Plato: Complete Works* (D. Zeyl, Trans., pp. 1224-1292). Cambridge: Hackett Publishing Company, Inc.

Rescher, N. (2000). *Kant and the Reach of Reason.* Cambridge: Cambridge University Press.

Rodl, S. (2014). Review: Transcendental Philosophy and Naturalism. *European Journal of Philosopy*, 483-504.

Ryle, G. (2000). *The Concept of Mind.* London: Penguin Group.

Schelling, F. (2006). *Philosophical Investigations into the Essence of Human Freedom.* (J. Love, & J. Schmidt, Trans.) New York: State University of New York Press.

Sextus-Empiricus. (2000). *Outlines of Scepticism.* (J. Annas, & J. Barnes, Trans.) Cambridge: Cambridge University Press.

Smith, N. K. (1913, October). The Meaning of Kant's Copernican Analogy. *Mind*, 549-551.

Taylor, C. (1989). *Sources of the Self.* Cambridge: Cambridge University Press.

Tolstoy, L. (1993). *War and Peace.* (L. Maude, & A. Maude, Trans.) Hertfordshire: Wordsworth Editions Limited.

Whitehead, A. (1985). *Process and Reality.* (D. Griffin, & D. Sherburne, Eds.) New York: The Free Press.

Williams, B. (1981). Persons, character and morality. In B. Williams, *Moral Luck: Philosophical Papers 1973-1980* (pp. 1-19). Cambridge: Cambridge University Press.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus.* (C. K. Ogden, Trans.) London: Routledge & Kegan Paul.