

Per Inge Mathisen

Moral Constructivism and the Appeal to Autonomy

An investigation into the moral philosophy of Christine
Korsgaard

Master's thesis

Trondheim, March 2007

Norwegian University of Science and Technology
Faculty of History and Philosophy
Department of Philosophy

Academic supervisor: Truls Wyller



Table of Contents

Introduction.....	2
Part I - Exposition.....	7
1. Voluntarism.....	8
2. Realism.....	12
3. Reflective endorsement.....	14
4. The Appeal to Autonomy.....	17
5. Self-identity and the unconditioned.....	23
6. Morality and evil circumstances.....	28
7. The moral interest.....	31
Part II - Analysis.....	34
8. Three conceptions of autonomy.....	35
9. Receptivity and agency.....	39
10. Practical and theoretical reason.....	45
11. Evil circumstances and the Formula of Universal Law.....	54
12. Practical reason and the fully reflective person.....	56
13. Revisiting the unconditioned.....	59
Part III - Conclusion.....	64
14. Synopsis of Korsgaard's argument.....	65
15. Reflective realism.....	68
Bibliography.....	71
Notes.....	73

Formulas and propositions:

Formula of Universal Law:

“Act only in accordance with that maxim through which you can at the same time will that it become a universal law.”

Formula of Humanity (as an End in Itself):

“So act that you use humanity, whether in your own person or that of another, always at the same time as an end, never merely as a means.”¹

Introduction

We cannot see moral values in the microscope, telescope, particle accelerator or mathematics, or in any other tool used to understand the physical world. And yet, we see the world as being full of moral values and ideals, and we discuss them as if they were something tangible which we could find out something about.

In fact, moral values would appear to have the most peculiar properties, which are not seen in anything else. They seem to motivate us to act, and oblige us to do things we might not have much desire to do. When we debate moral values and learn them from others, we seem to presuppose that there exists an objective answer to moral questions, yet it is hard to point at what in the world it is that corresponds to this objective fact.

The philosopher John Mackie put the problem like this:

If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else.²

According to Mackie, therefore, objective moral values cannot be in the world, and since there is nothing that statements about moral values then could refer to, they would all have to be false. He is not the only philosopher to voice similar scepticism about the existence of moral values. Moral philosophy has been troubled by the question of morality's objectivity for as long as there has been moral philosophy.

The question has two aspects. The first is how moral precepts can be said to be objectively true for all people, as we often think they are. A theory of morality which says that our moral judgements express our belief about moral facts is called *cognitivist*, and a theory of morality where moral judgements are made true by virtue of their correlation with entities that are intrinsically normative is called *realist*.

The second aspect is whether and how morality is motivating, as we usually presuppose it is when we debate it. This aspect is no less troublesome than the first. It seems rather obvious that we sometimes are motivated by ethical concerns, and that when we are convinced that something is morally correct, we act on this conviction. And yet we sometimes ignore such concerns and do things that we “know” to be wrong. A theory of morality has to account for all of this.

One line of response to the challenge started with Immanuel Kant, and has in later years been called 'moral constructivism'. Philosophers such as John Rawls, Onora O'Neill, Henry Allison, Paul Guyer and Christine Korsgaard can be said to belong to this tradition, which has gone through a great deal of change and increased popularity since John Rawls' interpretation of Kant in *A Theory of Justice* and Onora O'Neill's *Acting on Principle* in 1975.³

The term 'constructivism' in moral philosophy became popularized by John Rawls, who called his moral philosophy a variant of 'Kantian constructivism'. The basic idea of moral constructivism is that there are moral facts and moral objectivity, like cognitivism and moral realism claim, and they are constituted in the world by us, and the only moral entities in the universe are those entities that are making moral judgements. That is, us. In this view, morality is not a project of *discovery* about how the world really is, but a *rendering consistent* of our thoughts and reasons. This makes moral constructivism a *procedural* rather than a *substantial* form of moral realism.

I will look at the attempt made by Christine Korsgaard to give an answer to these questions by putting emphasis on the role played by the practical concept of autonomy. She finds that the fact that we have values to be “the most striking fact about human life”, and throughout several publications, she attempts to find out what values are and how they can be endorsed.

The problem of values is more apparent today. Korsgaard writes that “during the transition from the ancient to the modern world a revolution has taken place - in the full sense of the word. The world has been turned upside down and inside out, and the problem of value has been the reverse of what it was before.” The reason for this is that in the ancient world, “value was more real than experienced fact”⁴. In the philosophical systems of Plato and Aristotle, everything in the world attempted to attain its perfect form, whose existence was at least as real as matter. Today, however, what is seen as most real in the world is matter, not form. Form is no longer something that the world strives to achieve, but “must be imposed on the world of matter”⁵.

A key question for Korsgaard is how we can justify our ideas of morality. We are social, historical, and passionate biological beings that recognize a multitude of demands. They range from the complex demands of tradition to the elementary demand of the stomach. We evaluate these demands all the time to see which we feel imperative, which are optional, and which ought to be discarded. A key question is how can we know that we evaluate them correctly? This is not a question of understanding our psychology or biology. As she puts it:

“People who take up the study of moral philosophy do not merely want to know why those

peculiar animals, human beings, think that they ought to do certain things. We want to know what, if anything, *we* really ought to do.”⁶

It is obvious that we sometimes use some normative demands to vindicate others, but if this is all we can use, it threatens us with vicious regress. What would vindicate the vindicating norms? If that is all there is to it, we would reason in a circular manner, using normative demands given to us in advance to evaluate other normative demands. This gives rise to the sceptical challenge, that morality is just a convention, an invention, or a conspiracy, and that once we understand how morality works, we realize that there is no such thing as ‘the good’, ‘right intentions’ or ‘good deeds’, but at most useful intentions and useful actions.

It is important here to distinguish between two rather different questions. The first is for a theoretical, third-person account of how morality works in human beings. This question can be answered with reference to psychology or neuroscience, and is, not very controversially, a question of empirical fact. The other question is the one each one of these human beings must ask him or herself: *What to do?* This practical, first-person question seems in many cases strikingly different from the former question, and this too, is a puzzling question. What makes the practical question at least *seem* to be so different from a theoretical third-person account?

This is a pivotal question for Korsgaard. According to Guyer, “what most distinguishes her work” from that done by other Kantian writers, “is the concentration of her attention on Kant's attempt to derive the fundamental principle of morality and its several formulations from the very concept of practical reason itself.”⁷ The existence and motivating power of values and morality is derived from the necessary requirements of practical reason and the first person account.

The solution to scepticism usually chosen by substantive moral realists is to say that the regress of justification must stop somewhere, and that there is a moral ground that is real and final. To go further than this, and ask what vindicates this ground in turn, is to ask a meaningless question. Our biological history or nature is often where the regress is asked to stop.

The constructivist tradition finds this unacceptable, and therefore holds that the mental faculty of reason and understanding, isolated from conditional and contingent factors must be used as a test for the validity of all normative demands. If we accepted normative demands without first endorsing them using reason alone it would mean acting on impulse and our will would be determined by outside forces.

When we choose to act on normative demands that we can endorse using reason alone, without

falling back on and consulting other normative demands, we make ourselves a free and autonomous agent of our own will. This should give us a morality that is not contingent, and which stops the regress of normative demands. Furthermore, some moral constructivists argue, this is the only way to be a person with a will, and everyone must desire a free will, and so it should give morality a secure foundation against the sceptical attack.

Korsgaard sets up four categories of theories of what makes morality normative. The first is voluntarism, in which “obligation derives from the command of someone who has legitimate authority over the moral agent and so can make laws for her.” The most famous proponent of this theory is Hobbes. The second is moral realism, in which “moral claims are normative if they are true, and true if there are intrinsically normative entities or facts which they correctly describe.” The third is reflective endorsement, in which “morality is grounded in human nature”, and where the pivotal question is, “all things considered, do we have a reason to accept” normative claims, where the test is that it is “good for us”? The fourth is appeal to autonomy, in which the “capacity for self-conscious reflection about our own actions confers on us a kind of authority over ourselves” that gives us both morality and moral interest.⁸

Although she is a proponent of the fourth category, they are all, according to Korsgaard, true to some extent, and the theory of autonomy depends on key propositions from the other categories:

“Voluntarists try to explain normativity in what is in some sense the most natural way: we are subject to laws, including the laws of morality, because we are subject to lawgivers. But when we ask why we should be subject to those lawgivers, an infinite regress threatens. Realists try to block that regress by postulating the existence of entities – objective values, reasons, or obligations – whose intrinsic normativity forbids further questioning. But why should we believe in these entities? In the end, it seems we will be prepared to assert that such entities exist only because – and only if – we are already confident that the claims of morality are justified. ... we must show that particular actions are right and particular ends are good. Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action.”⁹

In the first part of this essay, we will look at the four categories described above – voluntarism, realism, reflective endorsement, and the appeal to autonomy, and see how in Korsgaard's view, these lead up to her own view. Each category has its own chapter. The fifth chapter takes a look at the role played by the concept of self-identity and how the concept of the unconditioned can explain the transition from the categorical imperative to the moral law. The sixth chapter looks at her novel

approach to meeting some of the frequently voiced criticisms of Kantian views on morality by using Rawls' distinction between general and special conceptions of justice. In the seventh chapter, we will look at the crucial role given to our moral interest.

In the second part, I will make an assessment of the role played by the concept of autonomy in Kantian constructivism as advocated by Korsgaard. It is hoped that the criticism will be applicable to Kantian constructivism more generally, but no attempt will be made at showing that this is the case, as that would exceed the scope of this thesis.

In the eight chapter, we will look at three possible conceptions of autonomy that the Kantian can hold. Chapter nine concerns the concepts of receptivity and agency as they are used by Korsgaard. The tenth chapter will explore the role of inclination in forming our motivation. The eleventh chapter will consider whether Korsgaard's use of Rawls' distinction between general and special conceptions of justice really do solve the problems that the theory of morality encounters in circumstances that are far from ideal. This leads up to a discussion in chapter twelve about some important differences between Korsgaard's concept of practical reason and the idea of a fully reflective agent. In the last chapter of the second part, I will revisit the idea of the unconditioned and consider it up against Kant's critique of teleological judgement.

I will attempt to pull the threads of these discussions together in part three. In chapter fourteen, I will attempt a short synopsis of Korsgaard's argument and see which parts of the argument that I have found to hold up. In the final chapter I will outline a way that I believe the main objections that have been raised in the above discussions can be overcome.

Part I - Exposition

1. Voluntarism

Voluntarism arose historically as an answer to an old challenge. In the philosophy of history, it usually starts with Socrates. In a story familiar to any student of philosophy, he had a fallout with the sophists, relativists who often saw concepts like ‘justice’, ‘right’ and even ‘truth’ to be relative and without any universal meaning. The outlandish claim by his student Plato was that not only was there universal meaning to such concepts, held by everyone, which could not be reduced down to something else, but moral relativism and its adherents contradicted themselves.

The later humanist tradition, who embraced the same kind of *might makes right* relativism as the sophists before them, was however always in an intellectual confrontation with the scholastic tradition, whose intellectual roots were Christianity and the Greek philosophers, both containing powerful universalistic viewpoints.

The humanists “aim at justice as it appears from man’s standpoint”, in the words of Gentili. Their justice was relativistic and glorified the state’s right to make war whenever it pleased, emphasizing a *might makes right* thinking in relations between states. Richard Tuck traces this tradition very well in his work *The Rights of War and Peace*.¹⁰

Political thinkers could not easily afford to ignore Christianity and the arguments of the Greek philosophers. The humanists, as far as they were interested in philosophy, were also troubled. Rousseau saw the connection of universalism with Christianity very clearly:

“if one but looks back to very ancient times, one easily sees that the healthy ideas of natural right and of the common brotherhood [*fraternité*] of all men spread rather late and made such slow progress in the world that it was only Christianity which generalized them sufficiently.”¹¹

Christian thinkers long attempted to create a synthesis of the two views, usually under what we would label *voluntarism*. This meant that morality was made correct by someone's command and overwhelming power, but since this power was God's, morality could not be said to be relative to each person, and it was objective for us as human beings.

This, however, meant that morality was dependent on God's existence, and our ability to correctly figure out what his laws actually were. The latter turned out to be a question that often seemed oddly separate from the former, and eventually voluntarist theory did away with the requirement of a deity or a world already filled with normative meaning.

This was a change long in the making, and was put as a revolution in theory by Hugo Grotius and Thomas Hobbes. Both explicitly reject the idea of there being any normativity inherent in the world. It all originated with people. Now this was hardly a new claim, the humanists had been reiterating this point for a long time, as the sophists and the Roman orators before them. What was new, was the combination of such a rejection with a claim that there were still objective normative truths.

The synthesis Grotius first created was between the scepticism of Renaissance humanism, with their emphasis on civil law, and the Scholasticism heavily influenced by Plato, Aristotle and Aquinas, with their emphasis on natural law. This new foundation allowed him and later writers such as Hobbes, Rousseau and Kant to keep what they found useful in both traditions while avoiding some of the crucial problems each found in the other.

The influential writers of the humanist tradition of the Renaissance were often jurists, and, apart from Thomas Moore and Machiavelli, few are well known today. Like the scholastics, they were highly concerned with making sense of the key notions of civil law inherited from the Roman Empire. This concern was still vividly alive by the time Grotius started out, and Grotius frequently referred to Roman sources. Their notion of right and justice was, like the Roman, entirely civil and accepted no substantial notions of universally valid natural laws.

Grotius saw natural law as the source of individual rights which was “easily brought into systematic form” and “beyond question, so that no one can deny them without doing violence to himself.”¹² Schneewind writes: “it is through him that the idea of rights as natural attributes of individuals came to occupy a commanding place in modern European thought.”¹³

This signified a shift in interest in rights theories from the catholic to protestant writers. According to Schneewind, “Suarez was the last great exponent of the Catholic theory of natural law. Grotius was the first Protestant to claim the same conceptual space.”¹⁴

Pauline Westerman writes:

It is indeed true that Grotius thought that universal and eternally valid criteria and principles are grounded in nature ... But he unequivocally confines the meaning of nature to *human* nature... we see that they are all informed by the urge to self-preservation alone, to which reason and sociability are instrumental.¹⁵

It is here that Grotius sought to save the day, and combine the best of both traditions. According to Tuck, he subverted the concept of natural law as he “took the *jurisprudence* of war which had

developed among humanist lawyers, and derived a theory of individual rights from it”¹⁶, a theory of rights that used natural law ideas about rights as natural faculties from the scholastic writers.

The shift is away from nature as understood by earlier theorists, and towards the idea of the cognitive devices of contract and consent grounded in human reason. Suárez and others before him had used consent to *explain* how social institutions came into being, but for Grotius it forms the key role of normatively *justifying* these social institutions.

For Grotius and later Hobbes, morality could not be found in external nature, but in internal, human nature. As Hobbes put it, there is no “common Rule of Good and Evil, to be taken from the nature of the objects themselves.”¹⁷ The method to find truth in morality was by the construction of hypothetical social contracts to which nobody reasonable could disagree. The external world was just matter in motion, “which presseth the organ proper to each Sense” to impart upon us sense impressions.¹⁸

Later Samuel Pufendorf would try to reintroduce nature as a justificatory source into political philosophy. However, he also takes from Grotius and Hobbes the artificiality of society and of moral distinctions, and from Hobbes a deeply negative view of human nature. Pufendorf’s project seems a grand merger of natural law theory of the past with the consent theories of his day. “Aquinas had claimed that natural law served to *realise* the inclinations inherent in man’s nature. The break-down of teleology induced Suárez and Grotius to assert that natural law should point out which actions *agree* with human nature.”¹⁹

Pufendorf turned natural law from eternal and unchanging principles into the divine will of God as our superior legislator. Law only existed when there was a superior to impose it. Natural law as the will of God was an idea Aquinas also shared, but with Aquinas the will of God could be read from the teleology of nature. However, Pufendorf saw nature as the ‘motion of particles of matter’²⁰, and rejected teleology as so many in his day. So there was no way to read nature in the way Aquinas did. Westerman comments: “Pufendorf’s theory marks the end of natural law theory ... He emphatically discards the idea that nature, in whatever sense of the term, is or can give rise to a normative order. He equally denies the possibility for human beings to discover those principles by means of reason.”²¹

Korsgaard writes that Hobbes and Pufendorf, like the theologian voluntarists, did not do away with the requirement of a God for morality, but instead required it precisely because they “believed that it *takes* a God or a Godlike sovereign to impose moral properties on the indifferent world of

nature.” Unlike the theologians, however, they did not find the content of moral law to be the arbitrary decision of a legislator, they “thought that the *content* of morality is given by reason independently of the legislative will” and that the “the role of the legislator is to make what is *in any case* a good idea into *law*.”²²

She takes both writers to hold that moral action is that which “does the right thing because it is the right thing, because it is the law, and for no other reason”, and that “if sanctions were supposed to provide the motives for obeying moral laws, then moral laws would be mere counsels” because the reason why we should act on them is drawn from our own benefit. A good person, in their view, obeys the law because it is the law. On the other hand, if moral laws did not have sanctions, they would also only be counsels, because they would then not have “the special force of *requirement*.”

She points out that “according to this picture neither moral obligation nor its proper and characteristic motive, the motive of duty, is possible unless there is a legislator backed by the power of sanctions who can lay down the law.”²³ Furthermore, she writes:

“Hobbes and Pufendorf believed that the contents of morality is given by natural reason. What morality demands of us is what it is reasonable for us, at least as a group, to do. The rules of morality are the rules that make social life possible, and social life is necessary for human beings. And Hobbes and Pufendorf clearly supposed that in many cases this consideration could be motivationally sufficient as well. Pufendorf, especially, says that in the absence of obligation we could still do what is right because it is useful. The legislator is not invoked to supply the content of morality or even to explain why people are often motivated to do what is right. The legislator is necessary to make *obligation* possible, that is, to make morality normative.”²⁴

2. *Realism*

The voluntarist idea that obligation becomes possible because there is a legislator that can make what is in any case a good idea into a requirement depends on the availability of what Hobbes calls 'irresistible power'. In this case, you only get an obligation insofar as the legislator is capable of punishing you. If you can get away with a violation, you are no longer under obligation, it is then just counsel to you. As Korsgaard puts it, if "irresistible power is just power unsuccessfully resisted, then authority is nothing more than the successful exercise of power, and things always turn out right. For no one can ever do what he lacks the power to do. ... Then you will always do what you have reason to do, and you can do no wrong." If you have power to evade punishment, you will no longer be under obligation, and your actions will never be *wrong* any more, they could at most be bad for you.

The same consideration arises if we try to derive morality from some other natural power. She argues that Hume falls into this trap by seeing our sympathetic motives as source of morality. In this case, then "if you lack sympathetic motives, you lack obligations. Your obligations vary along with your motives, and so you can do no wrong." The conclusion, then, is a dilemma: "If we try to derive the authority of morality from some natural source of power, it will evaporate in our hands. If we try to derive it from some supposedly normative consideration, such as gratitude or contract, we must in turn explain why that consideration is normative, or where its authority comes from." According to Korsgaard, the response of the (substantive) realist is to "dig in his heels".²⁵

The realist answer is that morality is, after all, intrinsic to the world. Korsgaard comments wryly: "Having discovered that obligation cannot exist unless there are actions which it is necessary to do, the realist concludes that there are such actions, and that they are the very ones we have always thought were necessary, the traditional moral duties."²⁶ The argument is ended by fiat, and "conceives ethics as a branch of knowledge, knowledge of the normative part of the world."²⁷

As John Mackie pointed out in the introductory quote, there is good reason to be, at least initially, sceptical of the idea that there are moral entities in the world. If there were, they would be unlike anything else we know of, and we would need a way of knowing them unlike anything else that exists.

Korsgaard also criticizes those realists who hold that there are "certain natural human interests, like our interest in having pleasure and avoiding pain" which are inherently normative. This mistakes the task of normativity, she points out, since the "point is not to look for some sort of specifically

normative *object*, but to look *more objectively* at the apparently normative considerations that present themselves in experience. That you are, say, in pain, *seems* like a reason to change your condition; the question is whether it is one.”²⁸

Moore argued famously in his “open question argument” that whatever definition we give of the word 'good', we can still raise doubt about whether things that fall under it really are good, and this, Korsgaard argues, applies in full force to any natural human interest.

3. *Reflective endorsement*

Even if natural human interests and sentiments are not sufficient to establish normativity, they nevertheless may be seen to be a necessary precondition for it. As we saw above, circumstances may decide whether a sentiment turns out to be good or bad, and the test of correct circumstance is in every case our faculty of reasoning. What is needed in addition is a reason to see it as good. This, according to Korsgaard, is the view of John Stuart Mill, David Hume and Bernard Williams, and is what she calls 'reflective endorsement'. They find “that the answer to the normative question rests in whether those dispositions are ones we have reason to endorse.”²⁹

This may seem to merely lift the problem up one level of abstraction, because reason still needs a way to establish the goodness in each case that is to be judged. The mechanism that is deployed to solve this problem is *self-interest*. It is in our best interest to be moral, because in the long run, it makes us be more happy and feel better than we would otherwise. Korsgaard argues that this argument “is not circular. Morality provides a set of pleasures of its own,” and those who cheat on it lose out on them. Moreover, in Hume's view, we need the love and respect of others to feel love and respect for ourselves, and “even the knowledge that others *would* hate [a cheater] if they knew what he was up to will be enough to produce humility and self-hatred when he acts unjustly.”³⁰

As Hume put it:

“By continual and earnest pursuit of a character, a name, a reputation in the world, we bring our own deportment and conduct frequently in review, and consider how they appear in the eyes of those who approach and regard us. This constant habit of surveying ourselves as it were, in reflection, keeps alive all the sentiments of right and wrong, and begets, in noble natures, a certain reverence for themselves as well as others, which is the surest guardian of every virtue.”³¹

This line of thinking establishes, Korsgaard says, “the harmony of two potentially normative points of view, morality and self-interest.”³² Similar to the voluntarist view of Hobbes and Pufendorf described above, morality is not what you cannot get away with, as the sophists charged, but it is a good idea.

However, how come the Moore's open question argument cannot be applied to self-interest as well? Korsgaard attributes to Hume the opinion that the “moral sense approves of its own origins and so it approves of *itself*” as a way of short-circuiting such an objection. We would “approve of a

benevolence-approving moral sense, while we would deplore a malice-approving moral sense.”³³ It is interesting to see how this argument seeks to close off the possibility of a challenge. If the moral sense is what informs us of what is good, then we must also find the ability to find the things that we find to be good to also be good. There would seem to be no intelligible challenge that can be made against it. Since the moral sense and self-interest is, as it is claimed, congruent, they must then be good.

The most obvious objection to this line of argument is that the moral sense and self-interest may after all not turn out to be congruent. If this turns out to be the case, the argument fails, and Hume seems to concede this.³⁴ Yet how can we talk about a self-interest separate from our moral sense? If self-interest is looking after your own well-being, and well-being is what is good for you, and what is good is determined by your moral sense, surely they cannot be separate?

Korsgaard offers another objection. She offers the example of a good-hearted Humean lawyer for a rich client who has died, and who finds in the client's papers two wills, one that donates all the money to medical research, and a more recent one that gives all the money “to the client's worthless nephew, who will spend it all on beer and comic books.” Why not secretly shred the more recent will? “The lawyer believes that her disapproval of this action depends on the fact that actions of this kind usually have bad effects which this one does not have. ... Her own feeling of disapproval may seem to her to be, in this case, poorly grounded, and therefore in a sense irrational. And this may lead her to set it aside ... since I approve of just actions because they are, generally speaking, useful, why not simply do what will be useful?”³⁵

The lawyer turns into a utilitarian, just like Bentham turned into a utilitarian by reflecting on Hume's moral theory. The “lawyer does not believe that the claims her moral feelings make on her in this case are well-grounded. If she could cure herself of them then that is what she would do.” It is, Korsgaard argues, a case of “normative failure”.³⁶

The problem here is that the reflective endorsement test is used to see if one inclination is better than another by seeing whether it is more in line with the general workings of our moral sense. However, it would seem that this is not a reliable guide. In its particular responses, the moral sense can be mistaken. It is sometimes right to do that which feels wrong, and wrong to do that which feels right.

Does it help to consider the *general* workings of our moral sense? It would not seem so either. If we know the moral sense is wrong in some cases, how can we tell in *which* cases it is informing us

correctly? If we can use the concern of utility alone to override in some cases, why not let it work in all cases? This was, in effect, what Bentham asked.

Korsgaard argues that there is a way to rescue reflective endorsement from itself, and that is by going the opposite direction of Bentham. Instead of keeping self-interest and utility and dropping the rest as unnecessary baggage, she argues that Kant kept reflective endorsement but dropped self-interest and utility. For Kant, and for Korsgaard, reflective endorsement does not test for morality by testing our inclinations against something else, but whether, upon reflection, we can allow “the maxim of acting on it can be willed as a law.”³⁷ That is, we do not justify moral rules by comparing them to self-interest or our moral sense, but the process of finding of moral rules by reflection *is* morality itself, and all that is needed to judge them as moral is already present in reflection.

4. The Appeal to Autonomy

The appeal to autonomy is based on the lessons drawn from the three previous chapters. Korsgaard sums up:

“Voluntarists try to explain normativity in what is in some sense the most natural way: we are subject to laws, including the laws of morality, because we are subject to lawgivers. But when we ask why we should be subject to those lawgivers, an infinite regress threatens. Realists try to block that regress by postulating the existence of entities – objective values, reasons, or obligations – whose intrinsic normativity forbids further questioning. But why should we believe in these entities? In the end, it seems, we will be prepared to assert that such entities exist only because – and only if – we are already confident that the claims of morality are justified.

The reflective endorsement theorist tries a new tack. Morality is grounded in human nature. Obligations and values are projections of our own moral sentiments and dispositions. To say that these sentiments and dispositions are justified is not to say that they track the truth, but rather to say that they are good. We are the better for having them, for they perfect our social nature, and so promote our self-interest and our flourishing.

But ... it is not just our dispositions, but rather the particular motives and impulses that spring from them, that must seem to us to be normative ... we must show that particular actions are right and particular ends are good.”³⁸

The way to do this in 'appeal to autonomy' is to look at moral rules or laws as they are without looking at any other considerations. All these other considerations are considered “alien causes”, including our desires and inclinations. As Korsgaard puts it, the “free will must be entirely self-determining.”³⁹ Such a self-determining will is an autonomous will.

The appeal to autonomy can perhaps first be found in Rousseau, who writes in *Emile*: “Make me free by protecting me against those of my passions which do violence to me. Prevent me from being their slave; force me to be my own master and to obey not my senses but my reason.” He makes a distinction between being good and virtuous, in that “he who is only good remains so only as long as he takes pleasure in being so” while the virtuous “is he who knows how to conquer his affections; for then he follows reason and his conscience; he does his duty”.⁴⁰ Kant is, of course, the most famous proponent of the 'appeal to autonomy' view, but I will, where not otherwise explicitly

stated, consider for the purposes of this discussion his views to be the same as those of Korsgaard.

The argument, according to Korsgaard, is

- (1) “that we must act 'under the idea' that we have free will, where a free will is one which is not determined in accordance with any law external to itself;
- (2) that a free will, if it is to be a *will* at all, must nevertheless be determined in accordance with some law or other;
- (3) that it must therefore be determined in accordance with its own law – that is, be autonomous; and
- (4) that this shows that the categorical imperative is the law of the free will.”

Although she finds this an “astonishingly simple argument”, it raises a host of questions. Korsgaard readily admits that it is “impossible not to feel that some sort of sleight of hand has taken place”.⁴¹

Why do we have to act in accordance with “some law or other”? This seems to fly in the face of the seemingly intuitive idea that we can act without committing ourselves to a rule of any kind, but just do it. Korsgaard concedes that this can in some cases be the case, but it is a peculiar fact that those acts we find important, unlike mere habits, are those we do for reasons. She holds that we cannot even conceive of deliberate acts which are done without any reason at all.⁴² Reasons “are derived from principles”, or laws, because, according to Korsgaard, the will is a causality, and a causality must operate according to a law.⁴³

Moral judgement is for Korsgaard a rational enterprise, and must therefore be subject to constraints of consistency. If right and wrong are thought to be something more a mere opinion, and if morality was not to be a system of principles, rules, or laws, then how would we say that one thing is right and another wrong? If we are to be able to separate right actions from wrong, we must have some knowledge of the regularities that separate the one from the other. Such regularities must take the form of principles, rules or laws, which are several names for the same thing.

The law is often a hypothetical imperative, of the form I ought to do X, because I want to achieve Y. If we are to achieve Y, we usually have to persist in our activity X. For example, if I am to finish this paper, I have to lay aside other concerns and actions in order to concentrate on it. I therefore make it a rule that writing this thesis is more important than these other things, and I *ought* to do that instead of the other.

To the extent that I “constitute *myself* as the cause of an action, then I must be able to distinguish between *my* causing the action and some desire or impulse that is 'in me' causing my body to act.” This has significant consequences, because “if *all* of my decisions were particular and anomalous, there would be no identifiable difference between *my acting* and *as assortment of first-order impulses being causally effective in or through my body*.”⁴⁴ No consistent acting according to a set goal would be possible in this case. “If I change my mind and my will every time I have a new impulse, then I don't really have an active mind or a will at all – I am just a kind of location where these impulses are at play.”⁴⁵

Actively willing a choice involves having a reason for it that is more than the incentive or inclination to it that is presented to you. If an act is done entirely on the presented incentive without making any rule out of it, this would eradicate “the distinction between a person and the incentives on which he acts.” There would be nobody there to do the acting, and the person would be a mere automata.⁴⁶

This same theme is repeated in the first requirement. If indeed “the will allows its choices to be directed by an external force, it surrenders its freedom.” Our will is seen as our ability to stick to imperatives we have set for ourselves, and avoid being directed by any and all impulse that comes our way. Autonomy is then being able to choose which impulse to act on without basing that judgement on any impulse.⁴⁷

Acting on the rule that you will take the first or strongest impulse that hits you as sufficient reason to act, is what Kant labelled acting “heteronomously”. Such acts can also be law-like, as sometimes “when you act heteronomously, the law is imposed on you by means of a sanction” from an external agent.⁴⁸

This leads us to requirement three, that the law of a person with a free will must be a law it gives itself independent of outside impulses. This points to a faculty independent of external impulses, and which can evaluate them and choose between them. This is the faculty of practical reason.

Korsgaard holds that reason is split in two. On the one hand we have theoretical reason, which we employ to understand the world, and on the other, we have practical reason, which we employ to understand what to do. As Korsgaard puts it “we stand in two very different relations to our actions: we must try to understand them, but we must also decide which ones to do.”⁴⁹

This distinction is linked to the worlds of 'phenomena' and 'noumena'. The world of phenomena is the sensed world, in which we read the laws of nature and endless chains of causality. The world of

noumena is the world as it is 'in itself', which is not sensible, only intelligible, that is, it is conceived by the mind alone. This follows from the fact that everyone “distinguishes between things as they appear and things as they are in themselves. And everyone can apply this distinction to himself as well as to other things.”⁵⁰ We view ourselves as existing in both of these worlds, but for different purposes.

The problem here is that of free will. We cannot act in the world unless we can assume that we act with a free will. It is “not that you must *believe* that you are free, but that you must choose *as if* you were free.” Korsgaard illustrates this with a thought experiment:

“imagine that you are participating in a scientific experiment, and you know that today your every move is programmed by an electronic device implanted in your brain ... it will determine what you think. Perhaps you get up and decide to spend the morning working. You no sooner make the decision than it occurs to you that it must have been programmed. We may imagine that in a spirit of rebellion you then decide to skip work and go shopping. And then it occurs to you that *that* must have been programmed. The important point here is that efforts to second guess the device cannot help you decide what to do. They can only prevent you from making any decision. In order to *do* anything, you must simply ignore the fact that you are programmed, and decide what to do – just as if you were free.”⁵¹

So in order to act, you must step out of the theoretical framework, and see yourself as existing in a merely conceivable world, that of noumena, in which you are not determined by outside forces. This is not a metaphysical point about there being two different existences in the world. “The idea of intelligible causality is a practical conception, and our belief in it is an article of practical faith. It is not supposed to be theoretically employed, and it cannot be used to explain anything that happens.”⁵² Indeed, Korsgaard holds that, at least as far as moral philosophy is concerned, the question of whether we actually do have a free will is irrelevant.⁵³

These two ways of looking at the world are not only distinct, however:

“Reason has two employments, theoretical and practical. We view ourselves as phenomena when we take on the theoretical task of describing and explaining our behaviour; we view ourselves as noumena when our practical task is one of deciding what to do. The two standpoints cannot be mixed because these two enterprises – explanation and decision – are mutually exclusive.”⁵⁴

This position tracks the Humean distinction between beliefs, which is similar to theoretical reason,

and desire, similar to practical reason, very closely.

This bears on explanations of weakness of the will as well:

“From the explanatory standpoint of theoretical reason, nothing is easier to understand than that a human being might evade duty when it is in conflict with her heart's desire. From the normative standpoint of practical reason her sacrifice of her freedom for some mere object of inclination is completely unintelligible. These two standpoints give us two very different views of the world.”⁵⁵

The categorical imperative is determining our will in accordance with a law we make for ourselves. When we need to make a decision, “the will must have a law, but because the will is free, it must be its own law. And nothing determines what that law must be. *All that is has to be is a law.* ... The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it has the form of a law. And nothing determines what the law must be. *All that it has to be is a law.*”

Any law decided upon by a free will must be categorical, rather than hypothetical or conditional on something else. However, “the argument I just gave doesn't settle the question of the *domain* over which the law of the free will must range” and “it does not establish that *the moral law* is the law of a free will.” If the “law is the law of acting on the desire of the moment, then the agent will treat each desire as a reason, and her conduct will be that of a wanton.” And if merely “the law ranges over the agent's whole life, then the agent will be some sort of egoist”, because it “is only if the law ranges over every rational being that the resulting law will be the moral law.”⁵⁶

It remains to be shown that the moral law is the law that would be chosen by the agent insofar as she is rational. In showing this, we first we need to look at the concept of self-identity.

5. *Self-identity and the unconditioned*

The very reflective structure of the mind “forces us to have a *conception* of ourselves.”⁵⁷ This self-identity must work as a “mediation between concepts and conceptions” since a “view of what you ought to do is a view of who you are.”⁵⁸

Korsgaard uses this concept of self-identity to show how we come to have obligations to others and reasons for acting. Obligations come built-in to the identities we adopt, such as the obligation of a psychiatrist not to violate the confidence of her patients. “No ‘ought’ is needed here because the normativity is built right into the role.” She also considers the two well-known phrases “I couldn’t live with myself if I did that” and “Just who do you think you are” as clear indications that self-identity plays this mediating role. In fact, an “obligation always takes the form of a reaction against a threat of a loss of identity.”

For a person, the integrity of being able to live by his own standards and rules is “what makes him a person at all.” As we discussed above, someone unable to lay down rules for his conduct cannot be said to have any stable self. Such rules, Korsgaard claims, take the shape of a practical identity.⁵⁹

On this account, obligations do not have to be moral. The obligations coming from the role of a Mafioso are just as real as any other. However, such obligations, like most of the self-conceptions which govern us, are contingent and each of them may be shed. What we cannot shed is our “identity simply as a *human being*, a reflective animal who needs reasons to act and to live.” This means that we need *some* identity, *some* self-conception, or “you will lose your grip on yourself as having any reason to do one thing rather than another”.⁶⁰

Our human identity is not seen as absolute. Korsgaard writes that there “is no obvious reason why your relationship to humanity at large should always matter more to you than your relationship to some particular person”. However, it is to treat our human identity, as one of that kind, which is to have *moral* obligations and moral identity on her account.⁶¹

What remains to be shown, then, is why we should be *moral* in this sense and what is wrong with the obligations of a Mafioso. Why is it that we need to value our identity as human more than, say, an identity as an assassin, or the humanity in others, instead of just valuing it in ourselves? We need a way to establish firmly the necessity and the necessary *scope* of universal laws, which has always been a particularly controversial and tricky part of any Kantian moral theory. For this, we turn to the concept of the unconditioned.

Korsgaard makes an explicit departure from Kant in making a clear distinction between the categorical imperative, which is “acting only on maxims you can will to be laws”, on the one hand, and moral laws, which are “maxims that all rational beings could agree to act on together in a workable cooperative system”. The arguments of the preceding chapter sought to establish that the categorical imperative is the law of a free will, but it did show that the moral law is. This is a “question of the *domain* over which the law of the free will must range. ... It is only if the law ranges over every rational being that the resulting law will be the moral law.”

There are many other possibilities for a categorical imperative, than to be a moral law. For example, a person could make it a law for herself to only act on reasons that benefit herself, or to only act on reasons that benefit people of her own nationality.

Practical reason is a natural faculty, and has a natural way of proceeding in its deliberation. Crucial to this deliberate process, Korsgaard finds the concept of the 'unconditioned'. It is worth quoting Korsgaard at length to bring this out:

“A thing has conditional value if its value depends on whether certain conditions hold. For instance, the value of the means depends on the value of the end it serves; ... Even happiness is not valuable in all cases, and so is conditional. A thing has unconditional value if it has its value in itself and so has it under all conditions. Ultimately all value must spring from a source which is unconditionally valuable, for as long as we can question the value of something, we have not reached the end of its conditions. ... Since it is the objects of our own choices which we take to be good, and those objects do not have value in themselves, the sources of value must be something that rests in us. It is not our needs and desires, for those are not always good.”⁶²

It is important to see how this is not, on Korsgaard's account, merely a theoretical standpoint:

“The claim that reason seeks the unconditioned is not based on an analysis of the abstract concept of reason. It is more a claim about the plight of self-conscious beings who because we are self-conscious need reasons to believe and to act. When we go looking for those reason we find ourselves – via a form of regress argument that is perfectly natural to any rational being – on a road that leads to the unconditioned, a road that threatens to have no satisfactory stopping point.”⁶³

The justificatory regress will only stop when we find something that is not conditional on anything else. The only satisfactory stopping point “must, therefore, be our humanity, our rational nature and

capacity for rational choice”⁶⁴ and “the value of humanity *is* the unconditioned condition of all other value.”⁶⁵

It is “because we are human that must act in the light of practical conceptions of our identity, and this means that their importance is partly derived from the importance of being human.” That in turn, means that nothing can be more important for us than this self-conception, and “all value depends on the value of humanity.” As Hume said, we must endorse the grounds on which we find something good before we can find it to be good. As Korsgaard points out, this is also similar to Kant's “Formula of Humanity”.⁶⁶

This identity as human is not a contingent conception of our identity, Korsgaard claims. It is not something “you have constructed or chosen for yourself, or could conceivably reject.” If it is our identity as human that is the most important, then “this shows that human beings are valuable”, in plural. Being human is something we share, after all, with all other human beings.⁶⁷

Why is this the case? The reason is that if we are to value anything at all, we must value ourselves as rational beings.

When we consider a normative question, we go through several candidate reasons. If we think about them, try to ascertain their weight, we need to consider where they get their value from, or what their value is conditional on. We then find another reason why the first reason is valuable, and that reason in turn is conditionally valuable on another reason; and so the regress goes on. If we are uncertain enough, we may continue this regress until we come to a reason that is not conditional on another reason, and which can bring the uncertainty to a halt. This unconditional reason is on this account our ability to reason. If the candidate reason threatens it, we are obligated to disregard it as a reason.

However, each and every one of us should be able to realize that we are not the only rational being around. In fact, the ability to reason is apparent not only in us, but in other beings as well, and if “you view yourself as having a value-conferring status in virtue of your power of rational choice, you must view anyone who has the power of rational choice as having, in virtue of that power, a value-conferring status.”⁶⁸ This creates obligations to other people, and gives rise to the moral law.

When we carry out this kind of regress, we realize that “lawmaking is the activity by which we create values, and that every act of lawmaking expresses the value we place on ourselves.”⁶⁹

Now, as we saw earlier, we must act under the assumption that we are free and autonomous beings.

To sum up the argument so far:

“we regard some of our ends as good, even though they are obviously conditional; there must be a condition of their goodness, a source of their value; we regard them as good whenever they are chosen with full rational autonomy; so full rational autonomy itself is the source of their value. Since this holds for other rational beings as well as myself, I cannot act against their rational autonomy without violating my own; and so it turns out to be a good will that is the source of all value.”⁷⁰

This does not mean, of course, that this regress is immediately apparent to everyone:

“We seek the unconditional by imagining a person who reason all the way back, who never gives up until there is a completely undeniable, satisfying, unconditional answer to the question. Obviously human beings often stop reflecting very far short of that.”⁷¹

However, the same answer, the same reasons, are readily available to anyone with practical rationality, to any normally intellectually gifted human being. This is not a philosopher's theory, but how ordinary thinking is in fact carried out: “Philosophy is ordinary human reasoning rendered *persistent*.”⁷²

There is an additional point to be made here. Not only does she think that practical reason is shared by us all, but the reasons on which we act are not private, they are public.

Korsgaard employs the notion of 'shared reasons' which in her opinion comes naturally from our social human nature. She acknowledges that “modern moral philosophers have usually not considered it allowable to help themselves to this fact in arguments aimed at justifying morality”, since our “sociability seems to be too biological or contingent a fact to play a role in rational arguments.” However, she puts forward the idea that “if our social nature is deep, in the sense that it is the nature of our reasons that they are public and shareable, then justifications of morality can and should appeal to it.”⁷³

This is not merely an argument about humans being social animals that like living in packs, similar to hyenas or rats. Instead, it is more of an argument that “the space in which meanings and reasons exist – is a space that we occupy together.”⁷⁴

Korsgaard also points to agreements as an example of such shared reasons, because “when they can communicate, two different people can integrate their functions, and, for purposes of a given activity, become a single agent.” Two persons, groups, and states, can in her opinion all reason

collectively, like a single person, and reach shared conclusions.⁷⁵ This means, in effect, that “you can no more hear the cries of an animal as mere noise than you can the words of a person.” Words and cries present themselves to us as obligations, part of reasons we share.⁷⁶

Shared reasons features frequently in Korsgaard's works, with inspirations from a wide variety of sources. It is not, as I see it, a necessary part of the overall argument, but rather an additional argument to bolster the link between the categorical imperative and the moral law. I do not, however, find it one of her strongest, cohesive and best argued line of arguments. Therefore I have chosen not to go further into it in here, focusing instead on the more traditional Kantian elements of her work.

6. *Morality and evil circumstances*

One of the more controversial aspects of Kant's moral philosophy is what it demands of us in far from ideal circumstances. The example most often used for illustration is Kant's position on lying in the essay *On a supposed right to lie from philanthropy*. Here Kant writes that

if you have *by a lie* prevented someone just now bent on murder from committing the deed, then you are legally [auf rechtliche Art] accountable for all the consequences that might arise from it. ... if you had lied and said that he is not at home, and he has actually gone out (though you are not aware of it), so that the murderer encounters him while going away and perpetrates his deed on him, then you can by right [mit Recht] be prosecuted as the author of his death. ... one *who tells a lie*, however well disposed he may be, must be responsible for its consequences even before a civil court and must pay the penalty for them⁷⁷

In Kant's opinion, at least as shown in this essay, you may *never* lie, even to stop disastrous moral consequences. As Korsgaard comments: “Unsympathetic readers are inclined to take them as evidence of the horrifying conclusions to which Kant was led by his notion that the necessity in duty is rational necessity – as if Kant were clinging to a logical point in the teeth of moral decency.” Even his most sympathetic readers, however, find themselves at odds with Kant's conclusion, and strive to find some way to preserve his moral method while changing the conclusion, and Korsgaard is among these.⁷⁸

Korsgaard suggests that the Kantian moral theory is single-dimensional, and that it needs to be recast as a double-level theory, similar to that of John Rawls'. Rawls has a special conception of justice as an ideal, that is “not applied without regard to circumstances”, and a general conception of justice that “represents the point at which justice becomes uncompromising.” To show why we need a double level theory, she draws up the two alternatives that are possible for a single-level theory, to either regard moral duty as encompassing everything and without consideration to what we may want to do or not, or to regard us having a strictly limited sphere of responsibility outside of which we have no duty whatsoever. Neither possibility is very appealing.⁷⁹

The problem with a single-level theory that has a strict sphere of responsibility is that it “seems grotesque simply to say that I have done my part” when bad things happen and we could have stopped it, but had no duty to do it. It also fails to explain the phenomenon that we sometimes think we are justified to feel regret for having to do something that under the circumstances seem right, but that we would not have done under more ideal circumstances. Finally, it creates conclusions that

many consider entirely implausible when confronted with evil, such as Kant's above-mentioned and much-derided essay on the supposed right to lie.⁸⁰

The solution to this issue that Korsgaard proposes sets up a division between two formulations of Kant's moral law. The Formula of Universal Law reads:

“Act only in accordance with that maxim through which you can at the same time will that it become a universal law.”

Another formulation of the moral law, the Formula of Humanity, reads:

“So act that you use humanity, whether in your own person or that of another, always at the same time as an end, never merely as a means.”⁸¹

Korsgaard regards the former as the general conception of morality, which is uncompromising, and regards the others as giving the special conception of morality, which is “an ideal to live up to in daily life as well as a long-term political and moral goal for humanity.”⁸² This seems to me to solve a long-standing problem for Kantian moral theory, as indicated above. As concerns the right to lie to a murderer, it passes the Formula of Universal Law, because “the maxim of lying to a deceiver is universalizable.”⁸³ It would not pass the Formula of Humanity, but since it can be waived in concrete non-ideal circumstances, we can allow ourselves to do so to prevent a morally disastrous outcome.

This may also be answer to a problem that has been raised at Korsgaard by Michael Smith. In the article *Internal Reasons*, he delineates two different approaches to how we look at acting on reasons when we are informed by ideal conditions. One approach is the advice model, where the desirability of an agent doing something depends on whether his ideal self would want him to do it under the present circumstances. The other approach is the example model, where the desirability of an agent doing it would depend on whether his ideal self would want to do it under ideal conditions. Smith claims that Korsgaard follows the example model, and argues that this runs counter to what we should do in evil circumstances.

The example Smith uses to show this is of a person who is defeated in a game of squash, and becomes terribly angry and consumed by a desire to smash the opponent in the face with his racket. It is customary to walk over and congratulate the winner in good sportsmanship, but this would increase the risk of the anger spilling over into action. Since we might suppose that this anger is irrational, it would not be present under ideal conditions. So if we follow the example model, what

this person has reason to do is to walk over and shake hands, but if we follow the advice model, what this person has reason to do is to leave as quickly as possible to cool down his temper.⁸⁴

The solution that Smith recommends leaves something to be desired, however. As Korsgaard points out, the double level theory explains that it “seems appropriate” that we feel “[r]egret for an action we would not do under ideal circumstances ... even if we have done what is clearly the right thing.”⁸⁵ This regret is inexplicable in Smith's advice model. So the example model offers us the special conception of justice, to which the Formula of Humanity applies. The advice model gives us the general conception of justice, to which the Formula of Universal Law applies. While we recognize that we ought to shake hands and show sportsmanship, to show respect for the other as a person, we absolutely cannot risk causing violence, and the double level model therefore more closely corresponds to our intuitions about the given example.

7. *The moral interest*

Korsgaard argues that when “we adopt (or come to wholeheartedly inhabit) a conception of practical identity, we also adopt a way of life and a set of projects, and the new desires which this brings in its wake. In some cases our conception of a contingent practical identity will give rise to new motives in a way that parallels the generation of the motive of duty by the thought of the categorical imperative.”⁸⁶

We need the assumption that the obligations, standards and norms imposed upon us by our contingently adopted identities are real in order to act upon them. If we did not find them to be real and worth acting on, that is, if we did not endorse them, we would not have any grounds to do *anything*. A person without *any* identity is therefore almost impossible to imagine. Such a person would not be a *person* because he would lack the *personality* that can only come from adopting one or more identities.

These obligations would have to be assumed to be real and endorsed as valid before we can act on them *even if they are not moral*. Korsgaard agrees that obligations derived from our identities must be real even if they are not moral. Now what happens, as it often does, when obligations derived from the necessary assumption of endorsement of our identity collide with the moral law? If autonomy is also a necessary assumption for our actions, then we need to sort out the overlapping set of actions which can be done according to both assumptions in order to act consistently with reason.

However, do we need the assumption of autonomy to get to the moral law? What has been explained so far indicates that a rational will would choose to act according to the moral law, but without autonomy as interest, why would the will want to be rational? Korsgaard sees this as a genuine problem. We need to explain “why a purely rational will would have the moral law as its first principle, it does not show us exactly why we should do so.” She holds that it is unclear why we think that “morality is supremely important” without adding the arguments of autonomy. She concedes that why this is so may be “rather difficult to see.” The problem is, she writes, is that we “are willing to grant the importance of the autonomy we express in moral conduct only because we already think that morality is supremely important. But it is still unclear why we think so.”⁸⁷ She refers to Kant's expression of the same problem:

“It must be freely admitted that a kind of circle comes to light here from which, as it seems, there is no way to escape. We take ourselves as free in the order of efficient causes in order

to think ourselves under moral laws in the order of ends; and we afterwards think ourselves as subject to these laws because we have ascribed to ourselves freedom of will”.⁸⁸

The necessity of grounding duty in an interest of the acting subject is perhaps more strongly expressed by Kant. In *The Metaphysics of Morals*, he wrote (my underscores):

“Every determination of choice proceeds *from the representation of a possible action to the deed through* the feeling of pleasure or displeasure, taking an interest in the action or its effect.”

Furthermore, Kant held that “any consciousness of obligation depends upon moral feeling to make us aware of the constraint present in the thought of duty”, and they are “*subjective* conditions of receptiveness to the concept of duty ... and it is by virtue of them that he can be put under obligation.”⁸⁹ Kant also touched upon this in the concluding pages of *Groundwork*:

In order for a sensibly affected rational being to will that for which reason alone prescribes the “ought,” it is admittedly required that his reason have the capacity to *induce a feeling of pleasure* or of delight in the fulfillment of duty, and thus there is required a causality of reason to determine sensibility in conformity with its principles.”⁹⁰

This moral interest is, according to Korsgaard, only satisfied by acting morally. This is the only way that we can respect the autonomy of ourselves and others. Aware that bringing in freedom of the will opens the door to many objections, she argues that “we may distinguish between the *concept* of X, formally or functionally defined, from a *conception* of X, materially or substantially defined. The Kantian *concept* of free will would be “a will which makes choices independently of all alien influences,” that is, a will which is negatively free. A positive *conception* of freedom would be a material account of what such a will would in fact choose.” The moral law is then “the unique positive conception of freedom.”⁹¹

As we saw above, Korsgaard does not see freedom as solvable by theoretical reason. However, for practical reason, the situation is different:

“if the moral law does indeed provide the positive conception of freedom, then we know how a person with a completely free will would act. Motivated by the idea of the higher vocation freedom gives us, we can act that way ourselves. But if we are able to act exactly as we would if we were free, under the influence of the idea of freedom, then we are free. Nothing is missing: the will ... chooses to act on the moral law for the sake of maintaining

its freedom; and we can do the same. *By acting morally, we can make ourselves free.*”⁹²

The claim that nothing is missing notwithstanding, this argument does leave some questions to be answered. In the next chapter, we will be taking a critical look at Korsgaard's moral philosophy by taking a closer look at these questions.

Part II - Analysis

8. *Three conceptions of autonomy*

In the article “How Kantian Must Kantian Constructivists Be?”, Evan Tiffany argues that Kantian moral philosophers must choose between a regulative and a constitutive conception of practical autonomy. At heart lies the “all-important question ... whether one should conceive of the autonomous will, not as a metaphysical claim about what we must actually be like, but as a regulative ideal that we necessarily impose on our valuations in our quest for the good.” A regulative conception,

“is a condition that reason imposes on thought and conduct, without presupposing that such principles are actually constitutive of the way the world is ... simply requiring that we think of ourselves under a certain description.”

From a constitutive conception, on the other hand, it would follow that “we must think of the relevant value-conferring capacity as instantiated in actual agents”⁹³. So on this view, if the claim of practical autonomy is true, we are among the first causes in nature, and can act independently of being determined by physical causes.

As we saw above, Korsgaard argues that the autonomy that is revealed by morality is an article of “practical faith” and “is not supposed to be theoretically employed”. And yet, she argues that we “*make ourselves free*” acting on morality and that this constitutes a moral *interest* for us to act morally. Evan Tiffany argues that this is an “attempt to tread a very thin line between a constitutive and a hypothetical understanding of the value-conferring will” which is “too thin” to work.⁹⁴

One way to illustrate the differences between the two conceptions is to use the distinction that we saw Smith use above between the world of ideal circumstances and the actual world with our present circumstances.

On the regulative approach, our autonomy is hypothetical and belongs in the ideal world. In this world, we have a free will and can be an independent cause of actions, and our inferences in this ideal world must take this assumption into account. This lends itself well to the example model, which informs us, as we saw Korsgaard wrote above, of “how a person with a completely free will would act”, allowing us to “act that way ourselves.” Being a hypothesis, the ideal world cannot prove anything is the case in the actual world, but the actual world can be used to disprove the hypothesis by showing that its presuppositions or implications are false. This corresponds to what we saw Korsgaard write about the reality of the free will not being derivable from our practical

assumption of it.

So acting in the actual world in accordance with the example set by the agent in the evaluating world can only disprove a hypothesis we form in it, not prove it is true for the actual world.

However, on the advice model, where we take not the conditions of our hypothetical ideal conditions into account but those of the actual world, we do not know if the actual agent has a free will, and so that cannot be part of the set of assumptions we make use of in forming normative conclusions. Hence, the Formula of Universal Law, which, as we saw above, Korsgaard sees as the most fundamental moral law and must be derived from an advice model, cannot follow from this conception of autonomy.⁹⁵ And how can an agent have a moral *interest* in autonomy if it is merely a hypothesis?

This leaves the door open for the constitutive interpretation of autonomy. The Kantian can answer that the point is that we must not let ourselves be determined by appearances, that is, outside causes, but find *something else* that we can ground our will in that is *not itself an appearance*. According to this argument, we must assume that our will is free and undetermined, and therefore in the realm of noumena. We can and do act on this free will, and, the proof of the pudding being in the eating, we therefore prove that we do indeed have a free will in the actual world.

On the constitutive approach, our autonomy belongs in the actual world. We not only act *as if* we had autonomy, we *have* autonomy. Kant's idea of transcendental freedom is sometimes interpreted this way, although it remains a highly controversial topic in Kant scholarship. I will, unsurprisingly, not attempt to solve this exegetical controversy here, merely show its consequences in regards to the present question, which is that on a constitutive conception, there is no problem to derive the Formula of Universal Law from the advice model.

When it is understood as making a point about the constitution of the real world, that is, the constitutive approach, there is a whole array of objections that can be mounted against it from the standpoint of theoretical reason. Phenomena that we observe may not be fully determined after all, as quantum physics suggests. Our first person assumptions may be faulty, and better explained by psychology, neurobiology or theoretical philosophy. And so on. According to Tiffany, one consequence is giving up on naturalism, to which the current philosophical audience is “increasingly committed”⁹⁶. I will not consider any of these arguments here, just note that the weight and number of objections against a constitutive approach seem daunting.

However, there is a general answer to all objections of this kind, and which I believe Kant may at

least have suggested. This answer is that what we are trying to find is not a theoretical understanding of the free will or of how the mind works, but to establish an *internal critique* of practical reason in order to establish not what we ought to believe, but *what we ought to do*. Having limited the question in this manner, the objections fail to apply. When we act on our assumption of autonomy, we prove to ourselves merely that the assumption was true within the limited horizon of practical reason and only insofar as we are concerned with what to do, not what to believe is the case about the world.

This approach carefully side-steps the two pitfalls of the constitutive and the regulative approaches, and as I see it constitutes a third approach, one that is not a strict logical deduction from the assumption of autonomy to the real world, but instead relies on apparent facts of moral psychology.

On this approach, we do not form the idea that we have a free will when we act because we deduct it from a prior assumption. Instead, by acting we make it appear to ourselves that we have a free will. The point of distinction from the constitutive approach is that this still proves nothing about what the *actual* world is like – we are *still* operating internally to practical reason. The point of distinction from the regulative approach is that autonomy does not operate as a regulative principle, instead it is taken as a basic psychological fact which we can only experience by acting in correspondence with morality.

This means we must assume autonomy also in the actual world, because autonomy exists there *to us* as a real and felt possibility. We may never know that it is real, but we nonetheless do experience its realness as if it were. Or we may say that the experience of autonomy is real, but it has nothing to do with the question whether we are free or determined according to theoretical reason, and is not supposed to be a representation of such freedom. This conception would claim that autonomy is real to the extent it is experienced as such, as one could claim about pain that it is real to the extent that it is experienced as painful. It means that we would harm our autonomy if we acted contrary to morality, because morality *is* ways of acting that affirm our freedom, and this thus establishes our moral interest in morality. We do not acquire this positive experience of freedom when we act contrary to morality.

I will call this conception of autonomy the 'emotional conception', since it depends on a certain moral folk psychology. This approach relies, however, also on a strict separation between practical and theoretical reason. We must remain isolated within practical reason until we have worked out the principles of *what* we have a duty to do, and only then can we employ theoretical reason to figure out *how*. Failure to keep the two apart would cause heteronomy. In the next chapter I look at

what such a moral psychology entails, and in the following chapter argue that such a total isolation of practical reason is not feasible.

9. *Receptivity and agency*

In the Kantian tradition, acting merely on inclination means acting without autonomy. Therefore we must act according to the moral law, and the moral is backed by the force of our conscience, which makes us feel guilt when we fail and respect when we observe it.

In a crucial footnote in *Groundwork*, Kant makes a premeditated strike against an objection he foresaw: That his talk of “respect” for the law indicated an inclination, and that as such, his moral law would be no less “heteronomous”, that is, based on inclination, than anything else. He writes:

“It could be objected that I only seek refuge, behind the word *respect*, in an obscure feeling, instead of distinctly resolving the question by means of a concept of reason. But though respect is a feeling, it is not one *received* by means of influence; it is, instead, a feeling *self-wrought* by means of a rational concept and therefore specifically different from all feelings of the first kind, which can be reduced to inclination and fear.”⁹⁷

Now the obvious question that arises is, can there be other feelings like this? In the *Metaphysics of Morals*, Kant does find a few more, such as moral feeling, conscience, and love of human beings, and he insists that they constitute a “vital force” and without such self-wrought feelings, our “humanity would dissolve (by chemical laws, as it were) into mere animality”.⁹⁸

These feelings are, however, seen as very different from those that are the source of our inclinations. They are brought about by reason, not taken as a reason for acting, and are therefore *active* (self-wrought) while the other kind of feeling is *passive* (received).⁹⁹ Why are feelings of the latter kind suspect, and why are the former kind not?

As Raymond Geuss notes in a critical response to Korsgaard, there is “a long tradition of criticizing Kant's ethics on this account – namely that as an ethical agent I see my own desires and inclinations as alien entities from which I must keep my distance – starting with Friedrich Schiller”¹⁰⁰.

Korsgaard also seems not to be blind to the fact that feelings can have an active character. At one point, when defending the notion that reasons are shared, she asks how, “if the painfulness of pain rested just in the character of the sensations, it could help to lie down?”¹⁰¹

There are two questions here. The first is whether we should talk of inclinations as essentially *passive* and what, if any, exceptions there are. The second is whether we should consider our inclinations as *alien* to us, unfit to determine our actions, and what, if any, exceptions there are. These two questions are deeply intertwined, since on the Kantian account, an inclination is *alien* if

and because it is *passive*. If we are to determine our actions, we cannot base our actions on something that we passively receive – our autonomy would be threatened.

According to the Kantian account, practical reason chooses among passive inclinations, picks one to act on, and if this choice is not determined by the influence of an inclination, this gives us a good reason to act and we act autonomously and without violating the moral law.

In response to the criticism from Raymond Geuss about why we should consider our inclinations alien to us, Korsgaard concedes:

“it is an important fact about human beings that our first-order desires and impulses do not all derive immediately from instinct, like those of other animals, but rather arise from a complex interplay of instinct and reason. ... Yet it cannot be denied that Kant seems to have made little use of this point in his moral philosophy, and it does raise some questions ... I don't know why exactly Kant didn't take up this question. But whatever the reason, my own employment of the notion of practical identity is in part intended to rectify the problem ... Part of my intention in invoking the concept of practical identity is to break down Kant's overly harsh, and even in his own terms over-simplified, division between natural impulses that do not belong to my proper self and rational impulses that do.”¹⁰²

She immediately guards this concession, however:

“Although I have just been suggesting that we do make an active contribution to our practical identities and the impulses that arise from them, it remains true that at the moment of action these impulses are the incentives, the passively confronted material upon which the active will operates, and not the agent or active will itself. ... The reason I must identify with my principle of choice when I act really has nothing to do with whether my first-order impulses seem totally alien to me or I regard them as my own productions. It is rather that at the moment of action I must identify with my principle of choice if I am to regard myself as the *agent* of the action at all.”¹⁰³

The consequence of this is that if we do act autonomously, those “beliefs and desires you have actively arrived at are more truly your own than those which have simply arisen in you”¹⁰⁴. Yet the question here is whether our inclinations stand in *need* of a special treatment to deliver us from passivity and make us autonomous, if they are not passive to begin with.

If we look at the issue as one of *amount* of reflection given to impulses for action, then no doubt an

extra round of reflection cannot hurt. However, there will be doubts as to the *necessity* of this extra round, and if our inclinations are not *alien* to us, the autonomy argument for moral action falls. Since the moral interest theory would crumble if this falls, it is surprising that Korsgaard gives it so little mention, especially given her amazement that Kant does the same.

The trouble begins with the ascribing of some feelings to our passive, sensual nature, and others to our active, cognitive nature. This assumes that feelings from our passive nature are *presented* for reflection as they are, and that there is such a thing “as they are” for feelings. I doubt both propositions. The caress of a lover gives rise to very different emotions than the exactly same physical stimuli from a rapist. To explain the pain of rejection without involving the faculty of reason is meaningless. It seems obvious that reason intervenes directly into our perception of these feelings to give it meaning even before they are presented for reflection.

In fact, it goes counter to a basic Kantian insight about perception. The things we perceive are not given to us, rather we structure the world around us in terms given by reason. We must make feelings meaningful to us *before* we can take them for reasons, and indeed, there is no question “how was the feeling before we interpreted it”, because it only *becomes* a feeling to us when we give it meaning.

When we feel a specific sensation, it is because that is how we think its cause is meant to be interpreted. We need to think of it this way to understand how the same physical stimuli can give a great variety of resulting feelings. That means that the representation of a feeling *is* the sensation of a feeling. If we fail to perceive a cut in the skin as a representation, it cannot hurt. That, I suppose, is why hypnosis works as an anaesthetic - it removes the link between a physical cause and any representation of it.

Take the man who has been sitting for too long in a sofa, and decides to go jogging. All his joints and muscles ache and protest, but he takes the pain as a *sign that* he really needed this and it will do him good. Contrast with the anxious person who feels unusual pain in the joints, and takes that as a *sign that* something may be very wrong. The physical stimuli may be identical, but the normative impulse to be evaluated will be very different, and the difference is an interpretation done by reason. So by the time the feeling presents itself to us, reason has already made a first judgement call as to what it is. The “pain in the joints” is not something we have access to apart from our interpretation of it given by reason.

Consider also a man who is working so hard for disaster relief that his body suffers. He knows that

a great deal depends on him and he feels a strong sense of obligation to the victims. Because of this he feels no pain. This is *not* the same as meaning that he *on reflection chooses* to ignore his pain. He is literally not in pain. It is not a subject of reflection on a conscious level, and so there is no conscious choice he could make to ignore it.

Korsgaard responds to concerns of this type as follows: “I think that in any reasonable theory of moral psychology there has to be room for something like a kind of passive input that is an original spur to action, what Kant calls an incentive. ... So an action is motivated by an incentive that is actively endorsed and chosen in light of some principle. It is an essential part of this picture that there should be some element of passive input.”¹⁰⁵

I do not see why there is any need to posit such a thing on the regulative and emotional conceptions of autonomy. Here we operate at all times internally to practical reason, and what we can know from the perspective of theoretical reason of a natural being being under the need for external causes are irrelevant. Such psychological reasons cannot play any role in such conceptions of autonomy. When we speak of the needs for a moral psychology in the Kantian sense, it must be a folk psychology which is wholly internal to practical reason.

On the constitutive conception the distinction between passive and active feelings is essential. As we saw above, Kant saw feelings as a required component of moral, or any, action. So if we cannot operate with such a distinction, all our actions will be determined in some way by our feelings, and we will never be autonomous. This, I believe, clearly speaks against a constitutive interpretation.

According to Susan Hurley, Kant overcame the problem of passivity in the theoretical domain. As she sees it: “Kant rejects the myth of the given, as the slogan has it: he rejects the idea that the content of perceptual experience can be taken as an unproblematic primitive, which the mind passively receives from the world, a matter of pure input.”¹⁰⁶

Our problem is that this myth may be rising its head again in Kant's moral philosophy. Kant attempts the same solution in practical as in theoretical philosophy – it is we as an active agent who adds to the world certain categories that we cannot explain without explaining our active contribution to the world. Yet if we extend this to our feelings, our inclinations, then they will no longer be merely passive.

Ralph Walker explains Kant's theoretical philosophy in this way: The “data of sensibility have to be ordered and classified, and this ordering is something we must do. The data themselves do not tell us how to do it, or even that we need to do it at all. ... The mind can perform this synthesis in a

number of different ways ... how it acts in classifying things is not dictated by the data.”¹⁰⁷ This is exactly how it seems to me that it would have to act in relation to our inclinations as well.

Korsgaard attempts to use personal identity against the charge of relying on alien inclinations, and this works quite well against the charge that contingent reasons cannot serve to justify actions. We do actively adopt many contingent facts into our lives, among these our various roles and identities, and not doing this “is the mark of a kind of immaturity”. However, this defence only works if 'alien' is taken to mean 'contingent', which is not the issue here. As she points out herself, “to get this right” we need to “distinguish our attitudes towards *contingency* from our attitudes towards *passivity*.”¹⁰⁸

If inclinations are not passive, then there is no longer an issue of autonomy at stake either. In *Kingdom of Ends*, Korsgaard considers the *reason* that Kant introduced the concept of autonomy. She writes that it “enters Kant's moral philosophy as the solution to a problem. The categorical imperative is not analytic, and disregarding its claims is therefore not inconsistent. Yet it is supposed to present us with a rational necessity.” She notes that “the connection between rationality and freedom” in fact “troubled Kant” and the arguments for it “are obscure and appear to be different” between his two major moral philosophical works. She argues later that the “point is not that you must *believe* that you are free, but that you must choose *as if* you were free. It is important to see that this is quite consistent with believing yourself to be fully determined.”¹⁰⁹

Korsgaard argues that autonomous acting “has nothing to do with whether my first-order impulses seem totally alien to me or I regard them as my own productions”. However, if we have to identify with a “principle of choice” when acting merely because this is a requirement for acting, then how come this principle holds any moral value? As Hurley points out: “Our sense of ourselves as autonomous is unavoidable, but it may be illusory. Just because we must think of ourselves as autonomous, it does not follow that we are, any more than it follows that we are unified subjects just because we must be able to think of ourselves as such.”¹¹⁰ We have no reason to act on what we consider an illusion.

If we cannot single out respect and other *active* feelings, those brought about by pure reason, as non-heteronomous, then either conscience loses its sanction and interest, or it loses its independence. If it loses its power of sanction, then we cannot use it to establish the *requirement* of the law of our will, in Hobbes' terms, and if it loses its interest, we do not find a moral interest. If, however, conscience loses its independence, then we fail to establish autonomy, as the sanction of conscience makes us heteronomous, and acting on our conscience becomes just as good as acting on

any inclination.

10. Practical and theoretical reason

To recap, we have seen above that the Kantian can employ three different conceptions of autonomy. The constitutive conception runs into problems with its distinction between passive and active feelings. But also the the regulative and the emotional conceptions of autonomy run into problems if a strict separation between theoretical and practical reason cannot be maintained. In this chapter, I would like to investigate whether such a clear separation is possible.

As we saw in the introduction, practical reason is derived from the first person view of the world, while theoretical reason operates on a higher level of abstraction in which the first person view is not necessarily privileged. Any explanation of morality from the perspective of theoretical reason can be met with the question: “But why should *I* want to be moral?” Even the old fire and brimstone argument needs to appeal to the first person account by asking: “Surely *you* would not *like* to burn in hell for all eternity?”

It would seem then that all morality would need to be connected to our deepest feelings, desires or passions in order to move us, in order to motivate us to be moral, indeed, to do anything at all. Hume is counted as the foremost exponent of this idea, and, as I have tried to show above, Kant did not take exception to it.

Korsgaard places herself in the same camp. She writes that “nature has constructed every living thing so as to preserve itself. ... self-love is what gives our inclinations standing – why they automatically count as candidate reasons.”¹¹¹ She finds the fact that we act on inclinations a biological fact of life. Although we cannot find our inclinations a sufficient reason to act, “if we deny normative standing to our natural impulses there will be no source of *content* for our reasons at all.”¹¹²

So how does reason fit into all this, do we not sometimes act from reason as opposed to mere feeling? Korsgaard finds questions like this erroneous, since acting “from reason is not opposed to acting from inclination; reason and inclination have different functional roles in action.”¹¹³ So in Korsgaard's picture, every act is based on an inclination, and needs to be endorsed by reason according to the moral law to be autonomous and consistent with reason itself.

It is not, however, yet clear what inclination is. Korsgaard writes very little about it, but I believe her position would be similar to that of Michael Smith. I will elucidate below on why I think so. Smith suggests in *The Moral Problem* that desire can be thought of as containing either 'phenomenological content' or 'propositional content'. In the former case, desire equals a raw feeling

or sense data. In the latter, desire would be a certain proposition about what we would like to see changed in the world. Smith suggests that it is only the latter form of inclination that it makes any sense to speak about in regards to moral motivation because we can be mistaken about our desires, and we often express desires in propositional form.¹¹⁴

In the previous chapter, I argued that it makes no sense to consider feelings as fundamentally separate from reason, because reason is the arbiter that interprets physical causes and tells us which feeling we should map to it, a process of course not under voluntary control. This means, I argued, that they are already a representation to us the moment we become aware of them. Smith must therefore be right to reject the 'phenomenological content' view.

On the conception of desire that Smith develops, desire has a unique *direction of fit* to the world. Desires “are states with which the world must fit”, while beliefs are conceptions that must fit with the world.¹¹⁵ We must adapt our beliefs depending on how the world is, but we try to adapt the world according to how our desires are.

As we saw above, Kant mentions only the passion of respect in a footnote in *Groundwork*, dismissing it as “self-wrought”. In the latter work, however, he dedicates an entire chapter to the passions, without which he thinks reason could not produce morality. In his late religious and anthropological works, inclinations are seen as natural and, perhaps, in a more positive light than in his early works.

In Kant's *The Metaphysics of Morals*, the word “inclination” is defined as a “habitual desire” (Begierde), which is by nature passive and a threat to autonomy if taken as sufficient reason. Kant seems to agree with Smith: “The *faculty of desire* is the faculty to be, by means of one's representations, the cause of the objects of these representations.”¹¹⁶ In his anthropological writings, Kant puts it slightly differently: “Desire (appetitus) is the self-definition of the power of a subject to imagine something in the future as an effect of such imagination. Habitual sensuous desire is called inclination.”¹¹⁷ In other words, desire operates by means of representations to “imagine something in the future” and to be the “cause of the objects of these representations”, that is, we try to make them become reality.

It is difficult to see which sense Korsgaard speaks of when she talks about 'inclination', but given that what Korsgaard is attempting to do is to elucidate a *Kantian* conception of motivation and inclination, I assume that she uses the concept of inclination in the way I suggested Kant does in *Metaphysics*, that is, as having propositional content. If inclinations on her account were to have

only phenomenological content, it is difficult to understand how we could *act on them* as she suggested we do in the beginning of this chapter.

However, the content of a desire is but one factor, and we frequently speak of our desires as having a certain *intensity* as well. What does this mean? Korsgaard takes the existence of this attribute as given, but what role it plays seems unclear. In *Kingdom* she writes that the “fact that we are able to act against our strongest desires reveals to us that we are free”.¹¹⁸ Later, she has a somewhat different account:

“It may be that what actually happens when you make a choice is that the strongest of your conflicting desires wins. But that is not the way you think of it when you deliberate. When you deliberate, it is as if there were something over and above all your desires, something that is *you*, and that *chooses* which one to act on. The idea that you choose among your conflicting desires, rather than just waiting to see which one wins, suggests that you have reasons for or against acting on them. And it is these reasons, rather than the desires themselves, which are expressive of your will. The strength of a desire may be counted *by you* as a reason for acting on it; but this is different from *its* simply winning.”¹¹⁹

On her account, then, the strength or intensity of an inclination *is separate*, at least in principle, from any reason we have for acting on it. Furthermore, an inclination, or desire, must, at least in principle, be independent of any reason for acting on it. Finally, while Korsgaard does not say this explicitly, I believe it follows that she must agree with Smith that desire and belief, although both contain propositional content, belong to entirely “different existences” and can, in principle, always be separated.¹²⁰

Smith explains the distinction:

“On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are assessable in terms of truth and falsehood, depending on whether or not they succeed in representing the world to be the way it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood. ... Crudely, our beliefs tell us how the world is, and thus how it has to be changed, so as to make it the way our desires tell us it is to be. An action is thus the product of these two distinct existences: a desire representing the way the world is to be and a belief telling us how the world has to be

changed so as to make it that way.”¹²¹

On this distinction between belief and desire as separate existences, it follows that “believing that, say, I should give to famine relief does not require that, other things being equal, I have a reason to give to famine relief.”¹²² This, Smith thinks, presents a problem, because it does not go over well with a common intuition about moral judgement, that when we express a moral judgement, we have a corresponding motivation and reason to act accordingly. This is what Smith calls the “practicality requirement”. The way Smith sets out to solve it should by now be familiar. He sets up a distinction between motivating reasons understood by “citing the complex of psychological states that produce the action” on the one hand, and the normative “considerations I actually take into account in deciding what to do” and that “from the agent's point of view, justify her actions”, on the other.¹²³

On the perspective of normative reason, Smith argues that we may desire something that we do not have rational justification to desire, and we may have rational justification to do something that we do not desire. If we are rational, however, we should desire what we have rational justification for desiring. That means, in turn, that “if an agent judges that an act is right then either she is motivated to act accordingly or she is irrational.”¹²⁴ Smith then defines irrational desires as those that are “*wholly and solely* the product of psychological compulsions, physical addictions, emotional disturbances and the like; to the extent that they wouldn't be had by someone in a non-depressed, non-addictive, non-emotionally disturbed state.”¹²⁵

This definition of irrationality may be somewhat different than the one we saw Korsgaard apply, which was to take an inclination or desire as sufficient reason to act. It may also be quite similar, all depending on how far you go with the concepts of compulsion, addiction and disturbance. Is my rationally unjustifiable desire for sugar drinks really caused by a psychological compulsion? If it is deemed a psychological compulsion merely because it is rationally unjustifiable, then we have a circular argument. I am unsure how Smith would respond.

However, do desires explain motivation? As we saw, Smith develops the idea that belief is connected to motivation by the way of correctly functioning rationality. He puts it as the rule: “If any agent judges that it is right for her to G in circumstances C, then either she is motivated to G in C or she is practically irrational.”¹²⁶ Korsgaard, too, sees this connection, and requires there to be inclinations on which we can act in order to motivate us.

But what does it mean to 'be motivated'? If it means that one will actually do it, barring physical or external obstructions, then this is obviously false. Although we can adjust C to post ex facto explain

why we did not in every case, making the rule trivially true and rather uninteresting. Alexander Miller charges that it is trivially true because anything can be inserted into the “practically irrational” part of the equation¹²⁷, although I am less convinced of this charge, since it still means that there is an additional need to provide explanation for why the link did not change into action, shifting the burden of evidence.

Even if we have both a desire and a belief that we should do G, it does not mean we are going to do G. As John Searle has argued, when a belief and a desire are thought as casually sufficient for an action, the action would seem compulsive, not rational. This would be the case of a heroin addict, who has an urge to take the drug and a belief that before him lies heroin, and cannot stop himself from taking it.¹²⁸ There are also many things we believe possible or even right and do desire, that we do not do. I believe it is right to say kind words to those who clean the floors at work and desire to do so, but when it comes to it there are always reasons I do not. I may believe it is right to be a more polite person, and desire I was, but I still will not be.

What seems to be missing is the fact that some moral considerations are more important than others, and these should be separated from the idea of “circumstances”. If I am too busy at work to be polite to cleaners, I will still be correct thinking that I should be and desire to be, although I will not be practically irrational. I just place greater emphasis on another consideration. So Smith's statement is missing something important.

This “something” is the priority we place on a consideration. Where does it come from? What if we would instead formulate Smith's statement as: If any agent believes that the *most* right thing she can do is for her to do G in circumstances C, she will be motivated to do G in C.

This would, at first, seem completely counter-intuitive, since we often think we know better what we should do than we what we actually do. However, consider the case of person A who is the closest to rescue another person B who is drowning. Imagine freezing time for a moment, to ask him a series of questions. A may say that it is his belief that the most right thing he could do would be to try to save B, since he has a good chance of success and is the only one who can do it; however, his fear of himself drowning, even if small, makes him not want to. He also, it turns out, wants to be a hero and wants to avoid the guilt of watching B die. So he has both belief in what is right to do, a belief that he would most likely succeed, and a desire to be a hero instead of a coward. The question must again be posed as a matter of priority, and what does it show priority of? We might ask him if he thinks it is right for him to act out of fear in this situation, and let us imagine he says “no”. We must ask ourselves whether this answer can truthfully be told. Does not his inaction

show that he does find it right to act out of fear in this case?

What else can moral philosophers make out of this example? Smith might say that A suffers from an understandable case of practical irrationality. But is it 'irrational' to be afraid of death? A commonsensical answer might be that he is lacking courage, but what is this quality 'courage' that is lacking other than a strong belief that he should help another person in dire need?

It is interesting to see how we implicitly assume what I outlined as an explanation when we assign blame in such cases. If I come too late to work every day, I might profess to my boss that I do believe it is important to be early and I do desire to be a good example in this regard, but I just cannot – staying in bed some more minutes is too comfortable. He would say, and I think he would be right in saying, that I *lack understanding* of how important getting to work on time is. He could yell at me: “Do you find it *more important* to be comfortable in your bed than at work?” And I would, rightly, hang my head in shame, for my practical lie had just been exposed. I did not believe getting at work early was more important.

Could it not be the case that I think it is more important to be comfortable when I just wake up, and think it is more important to be early at work when I get there? This would suggest I am in some sense practically irrational, unable to properly factor in my future goals in my present deliberations, as Smith would perhaps suggest. But why should I consider the needs of my future self as relevant for my present self? Surely this is contingent on a *belief* that it is right to do so, that it is right to forsake present short term pleasures for long term goals, and the *strength* of this belief would make the difference in such cases. If I do not have much belief that the future holds promise for me, then obviously it will be difficult to avoid short term pitfalls.

It has also been argued that we may hold mistaken beliefs about what we desire. In one of his examples, Smith writes about a person called John who says that he believes his greatest desire is to be a great musician. However, it is his mother who has “drummed into him the value of music”, and so when she dies, he loses all interest in music. He asks: “wouldn't it be plausible that [he] was just mistaken about what he originally wanted to do” and that the desire was false?¹²⁹

Yet what would this mistake consist in? On Smith's account, any rational action *requires* both a belief and a desire to give a motivating justification for it. So unless John had been acting irrationally, he *must* have had some desire to practice toward being a great musician. This desire may have been derivative off and conditional on some other desire, such as pleasing his mother, but what desire is not both derivative and conditional on something else? The fact that a desire is

derivative and conditional does not invalidate it. So it would seem that we can never be mistaken, in retrospect, about having a desire to do something if we know we did it.

In all of this, desire as a category of its own would seem to play at best an entirely subordinate or at worst unnecessary role. I desire to be a hero because I believe that to be a hero is good. If the belief went away, so would my desire. I desire to lie long in bed in the morning because I, obviously, believe the comfort is good. People who attack gays as sinners and perverts, but themselves secretly court gay prostitutes, will claim weakness of will as their defence, but the charge of 'hypocrite' does sit loosely. A 'hypocrite' is exactly one who professes to one set of beliefs, but expresses another in action, and although he may never be a liar by words, he would be one by his actions. This means that the statement "X believes Y is wrong" can be shown to be false if the statement "X does Y" is true at the same time and under the same circumstances.

One might object that all beliefs about the good and the right are, in fact, desires. That is, an opinion on how we would like the world to be. However, this might seem just another way of stating the same as I did above. If said that I *desire* that everyone be early to work each day, instead of saying I believe this to be right, the meaning of the sentence would be the same. However, if I then ended up realizing that I did not *desire* that everyone shows up early for work, I just desired that *everyone else* show up early, the meaning is not the same. The statement "A believes that X is right under circumstances Y" is not the same as "A desires that X be the case under circumstances Y", and I do not see how one could carry out such a reduction of beliefs about morality into desires without sacrificing moral rationalism and the idea that there are objective moral truths.

The belief/desire model is also said to explain problems of *wishful thinking*. It is frequently held that if we believe something to be true merely or partly because we desire it to be true, then we are deceiving ourselves. However, it is also held to be true that we should believe the best about people we do not know, even though we may live in a world full of bad people, as long as this is practically feasible. This is often called to hold people in good faith.

In many social contexts, it is expected that we show a great deal of good faith, and disregard suspicious evidence to the contrary. The man who refuses to believe his wife is unfaithful is an example of this. He has no way of knowing that his wife is not unfaithful, and chances are that life will turn up many things that will look suspicious to the inquisitive. The wife will be most displeased if she suspects that he is considering such beliefs, and therefore decides to drastically increase the bar on evidence found worthy of notice. And so it may well happen that the husband is the last to find out, but there is nothing inherently *irrational* about it.

It may be objected that this all happens on the level of politeness and manners. However, nothing horrifies us so much as to discover that beneath a veneer of respectability and politeness, there exists a very different person, for whom what is expressed is merely instrumental and what is truly believed is something else entirely. As human beings, we are very good at peering through such veils of politeness, and to find that someone who expresses good faith really holds bad faith, can be even worse than to be treated with bad faith outright. We can find someone refreshingly cynical or like the attention of suspicion, but it is rare indeed to find anything positive about being deceived.

What this tells us is that desires or inclinations do not explain motivation. For each example where we separate the belief and the desire, and say that we can hold the belief without the desire without being practically irrational, we can also see that we can hold *both* the belief and the desire, and *still* not have the necessary motivation to attempt the action without being practically irrational.

So if belief and desire lead “separate existences”, then the existence belonging to desire is entirely subordinate. The separation of motivating and normative reasons was introduced to solve the “problem” of motivation to act on moral beliefs, given that a desire was said to be necessary to induce motivation. This separation is then a solution looking for a problem to solve. It is, I will argue, far from intuitive on its own. We *do* care about the motivational reasons why we act as we do. This is one of the main reasons people are interested in psychology – not to understand others, but to understand ourself.

As Searle argues, we are quite often not satisfied with an explanation that diverges between its normative and the motivational reasons. He gives an example: The motivating reason is “Jones voted for Smith because Smith is an old drinking buddy”, while the normative reason given by Jones is that he was “justified in voting for Smith because Smith was the most intelligent candidate”. The reason given by Jones is, I think it is commonly thought, *false*, because it was not the reason that was *effectual*. In fact, most would see Jones as worthy of *blame* for not seeing the real reasons he had for voting as he did. So what is actually the case from motivating reason does *matter* in our evaluation of morality.¹³⁰

Is the first person view uniquely privileged, then? If it is true as Locke suggested of men, that we see “their actions to be the interpreters of their thoughts”¹³¹, then surely an external observer can sometimes just as accurately tell what someone is thinking as she is. Have we not frequently seen someone acting out an emotion, yet being unaware of it herself until someone told her of it? “Are you nervous?” we may say to someone who appears agitated, but this fact may well have eluded her, and even elicit an objection that we know to be false, even though we may well believe the

person is attempting to be truthful in her statement.

To conclude, the above discussion shows that it is not reasonable to expect a clear separation between practical and theoretical reason, since we *do* care how the world really is from a theoretical, third person standpoint when we determine what should motivate us. If morality is to motivate us, then, its premises had better not be shown to be illusory or indeterminate from that perspective. The regulative and emotional conceptions of autonomy give us no reason to believe that the autonomy we assume, on the regulative conception, or the autonomy we feel, on the emotional conception, is real and justified from the perspective of theoretical reason.

11. Evil circumstances and the Formula of Universal Law

We saw in the first part how Korsgaard avoids the disastrous consequences of a single level either-or moral theory by positing a double level theory, where the Formula of Universal Law is the point where morality becomes uncompromising, while the Formula of Humanity informs us of what goals we should aim for but not necessarily always should apply to the actual world. Korsgaard holds that we ought not, due to evil circumstances, always do what the Formula of Humanity tells us to do “regardless of possible disastrous results.”¹³²

This runs into a problem that Korsgaard has indicated elsewhere but does not bring up in this discussion. One of the reasons given in *Sources* for separating the categorical imperative and the moral law is that she agrees that the Formula of Universal Law, which is merely the categorical imperative universalized to every rational being and is the closest formulation of the moral law to the categorical imperative, does not give clear answers about moral questions. Problems with combining various activities means that we need to vary the scope of the law to fit circumstances:

“For instance, there is a well-known set of problems about coordination, in which it turns out that you cannot will a maxim as a universal law not because the action is wrong, but because some coordination problem would arise if everyone acted in the same way. ... Some people use that sort of thing as an argument against the Formula of Universal Law. I am rather inclined to think that it shows us an interesting way of thinking about the role of the Formula of Universal Law in thinking about moral issues. If you run into a coordination problem ... that is *morally* important fact. So I suppose that means that I think of the Formula of Universal Law as a heuristic guide in moral thinking. But the Formula of Humanity is more powerful; it is more likely actually to give us correct moral answers.”¹³³

Consider the issue of whether I can will the rule that I shall become a surgeon. Can I will the rule that everyone becomes a surgeon? Surely not. However, I can will that everyone who has sufficient grades and interest can become a surgeon, and since this includes me, I can will it without contradiction. Yet if we gut the Formula of Universal Law in this way, turning it into what Korsgaard calls the categorical imperative, then it seems to require additional support to set the scope at a moral boundary. Otherwise, racist or sexist scopes, for example, may “solve” problems of combination in ways we would not find morally acceptable in an uncompromising way.

Another problem is that if it is true that the Formula of Humanity “is more likely actually to give us correct moral answers”, and the Formula of Universal Law suffers from coordination problems, is it

the latter that is uncompromising? The obvious answer is that we run into fewer problems that way, but that will not suffice as an answer, since we need to find a way to answer this question that does not already assume that the pattern of results that we get is wrong. This is because on the Kantian account, it is the formulas that determine what is wrong and what is not. Another possible answer is that we first have to render our reasoning as a categorical imperative, sequentially, and so it follows that the Formula of Universal Law is closer at hand. Yet the categorical imperative does not give us moral content, and once we have made the jump to the moral law, there is no longer any reason given from sequence of reasoning to prefer the one over the other.

While the double level version of the Kantian theory as Korsgaard suggests has a lot going for it, it also poses a challenge for the good will. According to Korsgaard, the good will, which is acting according to the moral law, is “the source of all the good in the world” and is “the only thing whose value is in no way relative to its circumstances or results”.¹³⁴

The double-level theory means, however, that acting on the good will in evil circumstances may after all not be a good thing, as sometimes we must do that which we should not in more ideal circumstances.

Korsgaard approaches the question briefly in her discussion of the problem that we cannot know the consequences of our actions for freedom. She writes that “Kant tells us that one who honors the moral law cannot avoid thinking about what sort of world he would create under the guidance of practical reason”, and we discover the possibility “that our intentions and actions will make a real difference in the world, but that we will have no control over what sort of difference they make – because the consequences of our actions will not be what we intend. ... This possibility then makes the appeal of freedom seem like a fraud.”¹³⁵ Kant believes that this is what justifies our belief in God, a being who will guarantee that ultimately there is a connection between good actions and good results, but, he thinks, this we can only hope for and have faith in, we can have no knowledge of it.

I take it that Korsgaard's acceptance of a double-level theory must be seen as a rejection of Kant's religious optimism. In other words, we must work for such ideal conditions to become reality, but to act on the good will no matter what the consequences, as Kant suggests above, cannot be right. The upshot of this is that the good will must also be conditional on the right context.

12. Practical reason and the fully reflective person

Korsgaard argues that the “subject matter of morality is not what we should bring about, but how we should relate to one another”¹³⁶, and this sentiment and rejection of consequentialism is a cornerstone of any Kantian enterprise. It is a clear rejection of consequentialism in all its forms.

The point of morality, Korsgaard holds, is about the only thing we can control with certainty, and that is our intentions, and the only unconditionally good thing is therefore the *good will*. As we saw above, this may not hold under evil circumstances. I will here argue that it will also not hold under circumstances of imperfect knowledge.

The problem arises from the way she sees things in the world turn as conditionally good and in need of justification, and our ability to make them unconditionally good by the way of the good will. She writes that a “thing, then, can be said to be objectively good, either if it is unconditionally good or if it is conditionally good and the condition under which it is good is met.” That is, “when its condition is met in the sense that it is fully justified and the reasons for it are sufficient.” And since “all objective value must come from unconditioned value, the good will is the source of all the good in the world.”¹³⁷

If we commit an act that goes exactly as intended and that passes the filter of our understanding of the moral law, it must follow that we must think of the consequences of this action as being unconditionally good. Since being against the moral law would be the *definition* of immorality, we cannot both say that it passes the moral law *and* is morally bad. If we *can* do this, there must be a separate source of morality we can use to deduce this moral badness, and that is categorically ruled out by the theory. Yet that is exactly what I wish to suggest we can.

Things in the world must be good relative to the knowledge we have of their consequences.

Consider again the case of person A who sees a person B drowning and feels an obligation to save him. Let us assume he is now a perfectly rational moral agent, but without perfect information. He concludes that his felt obligation to attempt rescue is justifiable and therefore unconditionally and objectively good. However, just a moment later his friend C informs him that there are strong currents that would make such a rescue attempt suicidal and both would be killed. Suddenly, A reconsiders his justification, and concludes instead that his felt obligation is not justifiable, and therefore not unconditionally good. C, on the other hand, made the same conclusion all along.

This does not make any sense if a) we consider the unconditionally good to be objective and agent-

neutral, and b) we hold that only pure practical reason can *confer* the status of being unconditionally good. It seems clear that we can be *mistaken* about whether something in the world is unconditionally and objectively good. But this assumes that we are mistaken in relation to something else outside ourselves. This is a very different kind of mistake than the one where we would reason incorrectly or suffer from practical irrationality, since these mistakes are internal to the agent and to practical reason.

It is easy to see how this problem can be avoided. If we include full access to relevant knowledge as a precondition for something to be said to be morally good, as explicitly required, for example, by John Rawls' deliberative agent and Michael Smith's concept of a fully reflective person, there is no issue. On this view, being mistaken about matters of fact does not pose any problem for morality. Korsgaard sees such a position as fundamentally similar to her own:

“Kantian positions in general set a high value on reflection and are idealizing positions in the sense that moral concepts, as Kant defines them, are derived from the ideal of a fully reflective person. The fully reflective person is a corollary of Kant's idea of the unconditioned. We seek the unconditioned by imagining a person who reasons all the way back, who never gives up until there is a completely undeniable, satisfying, unconditional answer to the question. Obviously human beings often stop reflecting very far short of that.”¹³⁸

However, this solution comes with a price which I do not see that Korsgaard could easily incorporate. It is entirely crucial to Korsgaard's whole enterprise that morality not become a mere *theory* which we can decide to consult, a philosopher's theory, but actually exists at the point of practical deliberation as a potentially motivating force, our source of moral judgements¹³⁹. As Schneewind comments: “One of her central themes is that morality is through and through the product of practical, not of theoretical, reason.”¹⁴⁰ This fusion of a claim to correct moral philosophy with a simultaneous claim of correct moral psychology is what gives her theory its greatest strength, but it also exposes an additional Achilles' heel.

If we add the requirement of perfect knowledge in order to say that our inclinations are morally good, our moral judgements will always remain inconclusive and uncertain at the moment of action, since we never act with perfect knowledge. Worse still, it challenges the very idea of the good will being able to confer the status of being unconditional on our inclinations or other conditionally good things in the world.

In the moment before A is informed of the strong current, both A and C would rationally and with good will confer *mutually contradictory* statuses onto the obligation both feel to attempt a rescue. Since the good will is on Korsgaard's account the only way the conditionality of something good can be resolved, then if two persons with good will in the same situation can always in theory conclude differently, there can be no objective unconditional goodness in the world. All unconditional goodness must be agent-relative.

We can contrast with the way we can resolve propositions about facts about the world. Here we say that we check our proposition against how things really are, and objectivity can be said to be guaranteed by the fact that we live in the same world. However, on the constructivist approach, the world does *not* confer unconditional value, that is the job of the good will, that is, the correct application of practical reason.

It could now be objected that A and C before A is informed of the strong current are *not* in the same situation. A confers unconditional value not on the obligation felt to attempt rescue as such, but on the obligation felt to attempt rescue given what A knows about the situation. They do therefore not contradict each other. This would, however, make all moral judgements entirely situational, which again would endanger morality's objectivity.

Another possible response would be to say that it is not the *obligation* as an inclination and its consequences which is important, but the *rule* on which we act. The old rule will remain moral even if new information is added, since we then must select a new rule which we should have acted by, which again would be moral. This may be correct, but is beside the point being made. The problem does not lie in insulating the concept of the *good will* from criticism, but in proving the link between it and the objects on which it is supposed to confer unconditional value. It is not enough to show that the good will is unconditionally good, it must also be shown that it is the *one and only* source of unconditional value. This is what seems to fail here.

13. Revisiting the unconditioned

The idea that morality must have an interest is crucial in Kant's works on moral theory. I think Kant thought that morality must have an *interest* because he could not free himself from the Humean idea that we can only act based on *feeling*, as opposed to on reason alone. In his anthropological writings, Kant defines freedom to be an inclination, which is “linked with emotion”, and that through its birth cry a newborn “announces this claim to freedom (which no other animal knows).”¹⁴¹ We saw above how Kant used active feelings such as conscience as necessary conditions for morality. I think this is worth emphasising.

A key problem, I think, with the way they see reason is that, like Hume, Kant and Korsgaard assume that reason must be *reactive*. Impulses for action always come from outside, but by using reason to reflect on itself we achieve a sense of autonomy from these outside influences, and attain freedom from being determined by them. Korsgaard emphasises clearly this reactive nature of reason in a footnote in *Sources*, where she writes that it “is only in cases of reflective rejection that the impulse to act or refrain has to 'come from reason'.”¹⁴² We saw earlier how this categorical separation between inclination and reason cannot hold up. We also saw how the idea that we have passive inclinations that are the source of non-freedom makes little sense.

Yet if autonomy fails as an assumption of practical reason, as I have argued above, then only our contingent identities remain with the force of genuine obligation. Relativism would be right after all. I think this conclusion can be avoided.

In *Kingdom* Korsgaard has an argument that she calls the *Argument from Spontaneity*:

“This argument, which I will call the Argument from Spontaneity, shows that there are not really two choices, morality and self-love, on an equal footing. The will that makes the categorical imperative its law merely reaffirms its independence of everything except law in general. Its dependence on law in general is not a constraint, for that is just a consequence of the fact that it is a will. ... Adopting the maxim of self-love is ... an action for which there could be no reason. Thus ... the choice of the maxim of self-love over that of morality is unintelligible.”¹⁴³

In other words, moral action is the only meaningful action. To act contrary to or without considering morality is to act for a reason that cannot hold up and is therefore false. If this is indeed shown to be the case, do we really need a separate moral interest? I believe that Kant's work on

teleological reason may be of considerable interest here.

Kant speaks at great lengths about ends. There are *ends in themselves*, which are rational beings, such as man, who “is the *ultimate end* of nature, and the one in relation to whom all other natural things constitute a system of ends.”¹⁴⁴ Then there are the ends we set, ends for us, which are our goals and aspirations.

What seems to have gone less noticed is that Kant thinks that man not only has an ability to set ends, but also to find ends. These “found ends” are similar to Aristotelian formal causes, and they seem to exist for some purpose.

We can find mechanical causes in nature, but “if we desire to pursue the investigation of nature with diligent observation, be it only in its organized products, we cannot get rid of the necessity of adopting the conception of a design as basal.” In fact, “we do not *observe* the ends in nature as designed. We only *read* this conception *into* the facts as a guide to judgement in its reflection upon the products of nature.”¹⁴⁵

This is entirely necessary for our understanding of nature, as it “is utterly impossible for human reason, or for any finite reason qualitatively resembling ours, however much it may surpass it in degree, to hope to understand the generation even of a blade of grass from mere mechanical causes.”¹⁴⁶

This search for purposes is an a priori faculty (or aptitude) of the mind. “Therefore of all his ends in nature, we are left only with a formal, subjective condition, that, namely, of the aptitude for setting ends before himself at all, and independent of nature in his power of determining ends, of employing nature as a means in accordance with the maxims of his free ends generally.”¹⁴⁷ This faculty or aptitude for purposiveness is what would have led early man to consider a rock as something more than a rock, but as a tool or a weapon *meant* for a particular purpose, such as crushing bones to find bone marrow or for killing prey.

It searches unceasingly for purposes in everything we experience, and this leads us to ask for a purpose of life itself: “I require a final end, and it is only pure reason that *a priori* can supply this – for all ends in the world are empirically conditioned and can contain nothing that is absolutely good, but only what is good for this or that purpose regarded as contingent.”¹⁴⁸

This has significant religious implications, as this leads up to the moral proof of God’s existence: “It is the original moral bent of our nature, as a subjective principle, that will not let us be satisfied,

in our review of the world, with the finality which it derives through natural causes, but leads us to introduce into it an underlying supreme Cause governing nature according to moral laws.”¹⁴⁹

If, contrary to this, man finds an end other than reason itself, then “if he converts it into his entire end, renders him incapable of positing a final end for his own real existence and of harmonizing therewith.”¹⁵⁰

This means we have a natural aptitude for purposiveness, to find reasons, purposes and motivations everywhere we look, we *feel* an urge for such speculation, and it is hard to avoid holding it to be of a higher and more worthy motivation than any other. In religious contexts, the longing for purposiveness is what validates religion and therefore the most central motivation we can have, while most other feelings are apt to distract us from the contemplative activity necessary to look for purpose, and are therefore often called temptations.

We say that the purpose of the acorn is to become a tree, and that the purpose of lewd dancing on the disco floor is to attract the other sex. Everywhere we look, we find such purposes, and probably in too many places. It is quite clear from empirical sciences that to do so is an exercise of a faculty of the human mind, and there can be no doubt that we have a strong instinctive desire to use this faculty. This use is not merely reactive, but actively curious.

What Kant might be asking as a precondition for morality is whether the maxims of our actions “make sense”, that is have a purposiveness, apart from the other feelings. This might be seen as a rather trivial question, which can be answered in any number of ways, yet that is not really the case. The stripping away of the other emotions also strips away every other motivation that can be a *reason* for acting.

As we saw above, Kant insisted that our reasons ultimately are motivated by feelings, and I think he is correct. This must be understood correctly, however. As Michael Smith pointed out, there is a difference between what we can call *motivational reasons* and *normative reasons*, where the former is how we can theoretically understand what makes us tick, and the latter are reasons *for us* that we use in defending our actions to ourselves and others. The idea that reasons ultimately have their ground in feelings must be understood as a motivational reason, not a normative reason. It is only the faculty of reason that can give us normative reasons for acting. Feelings can motivate us, but they cannot by themselves give us sufficient normative reason to act, since their value to us are always conditional, and it is always the faculty of reason that resolves this conditionality.

To take the example of lying - why not lie when it can benefit me? The concept of ‘benefit’ loses all

meaning without the motivations that we have now stripped away, so what is left is “why not lie”. So we have to ask what the ‘purpose’ of lying is, and we have to look at lying generally, because the concept of lying is a general concept that we apply to a specific situation. Kant then points out that if everybody were to lie, we face a contradiction, and then we fail to find purpose. So if we cannot find purpose in acting from lying, we can only act from other motives. Acting from these other motives is then what Kant calls heteronomy.

Kant says, in effect, that if we look at the maxims of our actions and see that they are compatible with acting from the feeling of purposiveness alone then they are acceptable from a moral standpoint. This is the basis. The rest of the theory, including the categorical imperatives, can be built directly upon this.

A question is if this is sufficient grounds for morality. It clearly is not, nor does Kant think so, at least not in his latest work, in *The Metaphysics of Morality*. He identifies several feelings which he thinks are necessary for morality and which it makes no sense to say we are under duty to feel, because they are necessary for duty. These are conscience, moral sense, respect and philanthropic love. Kant attempts to defend his choice of these feelings as foundational in this sense by saying that they arise from rationality, and do not determine rationality, and therefore do not impede on our autonomy.

If we understand rationality as the mere reasoning of the mind, then, as we saw above, this argument must be rejected out of hand. All kinds of feelings arise from our reasoning, and can be postulated to be necessary for whatever moral theory we have also postulated and attempt to make plausible. Making this distinction is then a way of begging the question.

However, if we understand rationality as the aptitude for purposiveness, as I have suggested, then Kant’s distinction may make more sense. The respect that Kant speaks of is not any respect, but the respect we feel for purposiveness found. The pangs of conscience that is relevant here is not any feeling of insufficiency caused by a failure to follow a maxim, since there are evil maxims, but only those caused by not following maxims grounded only in purposiveness. We can reformulate this as saying that feelings that arise from considering maxims in the light of purpose alone do not cause a problem for the concept of autonomy.

This reading of Kant seems to make autonomy an empty and arbitrary choice of one feeling over another. The problem of autonomy would seem to depend entirely on the supposition that the aptitude for purposiveness is special among our feelings, and alone grounds for the autonomy of a

subject that sets ends for itself, while the other feelings make us an object that is determined, not determining.

Kant put it this way: “it is only the faculty of desire that can give the required point of reference – yet not that faculty which makes man dependent upon nature (through impulses of sense), that is, not that in respect of which the worth of his existence is dependent upon what he receives and enjoys. On the contrary, it is the worth which he alone can give to himself”¹⁵¹ It seems easy to dismiss it as wishful thinking, or worse, a result of a religious pietism. While this supposition was no doubt caused by both wishful thinking and a very Protestant pietism, it has more going for it than that.

There is no way to conceive of morality except by employing the feeling of purposiveness. Rules and principles cannot have meaning unless they have a purpose. We can, as Kant points out, feel empathy and happiness, just like animals do, but this is not covering what we mean by morality, which is, as Kant points out, inherently systematic and governed by rules. To see the feeling of purposiveness as special in regards to morality is a rather plausible claim.

Part III - Conclusion

14. Synopsis of Korsgaard's argument

The arguments of the Kantian tradition in moral philosophy are varied, and do not follow each other neatly as a logical deduction. Nevertheless, I believe the main points of Korsgaard's position can be summarized as follows:

- A) To any theoretical, third-person account of how morality works, we can ask whether it is right *for us* to act in this manner.
- B) Consequently, no theoretical, third-person account of morality can tell us what to do in practice.
- C) Conversely, we can only answer the question *what to do* by involving the practical, first-person standpoint of morality.
- D) Consequently, if moral claims can be true, they can only be true if they correctly describe what is right for us to do from a practical first person standpoint.
- E) When we act, we act on an inclination that we have chosen among candidate inclinations.
- F) We chose an inclination to act on by acting on reasons, which have the form of principles, or laws to us. These laws are passed by reason, and are backed by the sanction of conscience.
- G) We naturally seek to have valid reasons for acting, just like we naturally seek to have true beliefs.
- H) If the validity of a reason is conditional, we will seek to ascertain the validity of the reason it is conditional on.
- I) The validity of all reasons is conditional, except for the reasons for which reason alone is sufficient reason.
- J) We must assume that we have an autonomous will that acts for valid reasons.
- K) If we let an inclination determine our choice, our will is not autonomous.
- L) Consequently, we must choose to act on a law that does not have an inclination as its end. (The categorical imperative.)
- M) We must assume that everyone has an autonomous will.
- N) Consequently, we must choose to act on a law that respects the autonomy of all individuals. (The moral law.)

I believe points A to D above are correct, and E is correct insofar as we count reason itself as an inclination among others, but this I do not believe Korsgaard would agree to. Point E may be

correct, but as I argued above, it is far from obvious what inclinations are. As to F, I would agree, but I think Korsgaard makes too much out of arguments about sanction and the supposed lessons from voluntarism. I will not go into any detail about it here, but if the sanction of conscience is supposed to render obligation a requirement rather than counsel, we must explain why conscience is sometimes silent even in the face of the most cruel injustice, and how this does not invalidate the requirement.

I believe points G and H are trivially true, although it is important that Korsgaard concedes that H does not happen automatically. This brings up the question of what makes us seek this condition, and if this is itself a normative or even moral activity. I have tried to outline an answer to this in the previous chapter where I look at such a process of regress as a natural faculty and operation of the mind which we cannot do without, not even in trivial reasoning. As I have tried to show in the previous chapters, I believe that points H and I hold, but that J fails, and consequently points K, L, M and N do not follow from it.

This means that the moral law fails to be established. However, I believe there is another way to establish the moral law. The problem with the concept of our autonomy as an acting agent is that it is an assumption from which nothing follows. The idea of there being inclinations that *determine* us and make us lose our spontaneity and autonomy is wrong, because we must just as much assume we are autonomous if we act on them as when we do not. The arguments that are based on receptivity and a constitute conception of autonomy fail, as we saw above, and I do not see how arguments based on the subjective experience of freedom carry any normative weight. How can a feeling of freedom render an action moral any more than a feeling of pleasure or happiness?

Furthermore, we see that a separation between inclination and reason cannot be done cleanly, for two reasons. Reason is used throughout in the constitution of our inclinations, and reason itself must be seen as a natural feeling, as our curiosity and thirst for knowledge. We also saw how autonomy is not necessary to establish a moral interest, that is, a motivation for morality.

Korsgaard is correct to see reason alone as capable of reflecting on itself and finding itself sufficient and unconditionally good. If the argument from the previous chapter is correct, we do not have a problem seeing this as sufficient interest for morality. This leads us first to what she calls the Categorical Imperative. If we agree that acting on personal motives itself is not a sufficient reason, as I suggest in the previous chapter, then we also come to the Formula of Universal Law, but that, as we have seen, has its problems being the sole ground of the moral law.

So it would seem that abandoning the argument from autonomy in the way suggested, leaves us with a highly impoverished version of the moral law. We need the Formula of Humanity, and in the final chapter I wish to tentatively look at a way that this can be achieved without making any assumptions about our own autonomy being anything other than a quiet and implicit assumption that has no force.

15. *Reflective realism*

In the introduction we looked at John Mackie's famous 'argument from queerness', as it has become known. I would like to quote it again:

If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else.¹⁵²

Korsgaard replies almost poetically to this in the closing paragraph of *Sources*:

But Mackie is wrong and realism is right. Of course there are entities that meet these criteria. It's true that they are queer sort of entities, and that knowing them isn't like knowing anything else. But that doesn't mean that they don't exist. John Mackie must have been alone in his room with the Scientific World View when he wrote those words. For it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. They are people, and other animals.¹⁵³

The question is, what is it about other people, and, apparently, animals to some degree, that give us the special obligations of morality? And, if we assume that morality must look like what the Formula of Humanity tells us it is, how can it not involve a concept of autonomy?

I believe that what I have shown in chapter thirteen is to show how *values* can be found and where they come from. In order to get to *objective* values, such as those Mackie argues cannot exist and Korsgaard understands them, we need to show not only that everyone has the capacity, or faculty, to be *able* to arrive at the same values, but that, upon proper reflection, *must* arrive at the same values. This would then be a conclusion of a procedural realism, and a moral constructivism.

In the Kantian appeal to our autonomy, we confer value on ourselves and *by extension* everyone else. This means, however, and this is my final argument against this theory, that the value of other people is *conditional* on the value we place on ourselves, and not unconditional, as is claimed. So if we cannot act solely on inclination because it is a conditional good, why do we have to refrain from acting solely from respect for other people, who also only have conditional value?

The major strength of Kantian moral theory is that it is firmly embedded in an epistemological framework. It is an internal critique of the working of practical reason itself, that is, what we have

or have not any reason to do. If we follow the idea that reason itself is a natural faculty and an active inclination of the mind, and that only this faculty and its natural operation can be said to be unconditionally good, I believe we must look at the sources for its successful operation to find the obligations that our reasons must conform to.

Purposiveness is not just to know purposes, which would never satisfy us, but the restless search for *new* patterns and purposes, larger meanings and deeper connections. That is, curiosity is an integral part of our faculty of reason. However, anything we create, for any reason, we already understand the purpose of, so it cannot satisfy our reason directly, it must get its meaning from something else that we can investigate, such as other people's appreciation of it. Yet for other people to react in a way that creates meaning, they must be outside of our control, their actions must not be determined by us. This goes for anything we investigate for purpose – if we know, create or control the result, any investigation is pointless. This leads us to conclude that it is not our autonomy that is the condition for reason, but other people's. So the Formula of Humanity is correct after all. Not as a logical, that is theoretical, extension of our own autonomy, but directly as a requirement of our faculty of reason.

What I am suggesting, then, is that the source of value is not ourselves, but first and foremost *other points of view* and the meanings and sources of meaning that are expressed in them. Since we also hold a unique point of view, we would *by extension* also see ourselves as having moral value, but only a conditional moral value. This would explain, against the Kantian account, why we would not just sometimes sacrifice our life for those of others, but consider doing so *morally imperative*. The member of the resistance who jumps out a window to certain death to escape capture and torture which might expose his associates makes *moral* sense on this account. On the Kantian account, such actions are merely unintelligible, since they destroy the source of our reasons for moral action, that is, the acting agent.

I believe Korsgaard, in her attempts to render Kant consistent and interesting, have come closer to the point that I have outlined here, than Kant did, although I am sure she would disagree with my conclusions. The emphasis on our self-conception leads in this direction, as in her emphasis that we “live in the eyes of others because we must live in our own” and her description of us as beings that need confirmation is spot on:

“Self-conception is the source of some of our sweetest pleasures – knowing ourselves to be loved or to have done well; and our greatest (and often self-inflicted) torments – believing ourselves to be worthless, unlovely, or unlovable. ... A sense of personal worthlessness ...

[can be] the germ from which nihilism and the rejection of all value spreads.”¹⁵⁴

We usually act under the assumption that we have an audience. This imaginary observer is what gives our actions purpose. There are acts which have no purpose, such as habits, but to the extent our actions are thought to be purposeful, it must be purposeful for *someone*. To be someone is to adopt a practical identity, and the meanings of an identity are relational to other people. An identity cannot exist in isolation, it demarcates us from others.

Since what we create is already known, we do not create value, we only discover it, rework it, and reshape it. The more variety and richness we find in the world, the richer our normative universe becomes. To become richer in value, we must explore the unknown and restrain ourselves from the impulse to control our surroundings and our fellow human beings, so that they can develop in ways that are new to us. Therefore the first principle of morality must be restraint toward other people, as the Formula of Humanity says.

I believe, and here I must part company with the Kantian conclusion, that we also have an obligation toward the other sources of wonder and experience that we find in nature. That is, we have an obligation to preserve the richness of the world and the conditions that gave rise to our identity as a rational being in it. Although such sources of inspiration, wonder, learning and experience are immeasurably more distant to us than other people, or even animals, they are nevertheless a powerful factor in our search for purpose and meaning. What is, for instance, the allure of the starlit sky, but precisely its tantalizing suggestion of endless possibilities for discovery?

Bibliography

- Edmundson, William A. *An Introduction to Rights*. Cambridge University Press, 2004.
- Greene, Joshua. Cognitive Neuroscience and the Structure of the Mind. *The Innate Mind*. Oxford University Press, 2005.
- Hart, H L A. *The Concept of Law*. Oxford University Press, 1961.
- Hobbes, Thomas. *Leviathan*. London: Penguin Books, 1968 (1651).
- Hume, David. *A Treatise of Human Nature*. Oxford University Press, 1978.
- Hurley, Susan L. *Consciousness in Action*. Harvard University Press, 1998.
- Joyce, Richard. *The Evolution of Morality*. Cambridge: MIT Press, 2006.
- Kant, Immanuel. *Anthropology from a Pragmatic Point of View*. Southern Illinois, 1996.
- Kant, Immanuel. *Practical Philosophy*. Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason*. Cambridge University Press, 1998.
- Kant, Immanuel. *The Critique of Judgement*. Oxford: Clarendon Press, 1952.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge University Press, 1996.
- Korsgaard, Christine M. *Internalism and the Sources of Normativity*. First edition published by Stanford University Press, 2002. Second, cited edition only available at <http://www.people.fas.harvard.edu/~korsgaard/>.
- Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.
- Locke, John. *An Essay Concerning Human Understanding*. Dover Publications, 1959.
- Miller, Alexander. *An Introduction to Contemporary Metaethics*. Polity Press, 2003.
- Nichols, Shaun. Innateness and Moral Psychology. *The Innate Mind*. Oxford University Press, 2005.
- O'Neill, Onora. *Constructions of Reason*. Cambridge University Press, 1998.
- Rawls, John. Kantian Constructivism in Moral Theory. *Journal of Philosophy*, Vol. 77, No. 9
- Raz, Joseph (ed.). *The Practice of Value*. Oxford: Clarendon Press, 2003.
- Ryle, Gilbert. *The Concept of Mind*. London: Penguin Books, 1990 (1949).
- Schneewind, J B. *The Invention of Autonomy*. Cambridge University Press, 1998.
- Shafer-Landau, Russ. *Moral Realism – A Defence*. Oxford: Clarendon Press, 2003.
- Searle, John R. *Rationality in Action*. Cambridge: MIT Press, 2001.
- Smith, Michael. Internal Reasons. *Philosophical and Phenomenological Research*, Vol. 55, No. 1. (March, 1995).

- Smith, Michael. *The Moral Problem*. Oxford: Blackwell Publishing, 1994.
- Stanley, Jason. *Knowledge and Practical Interests*. Oxford University Press, 2005.
- Tiffany, Evan. How Kantian Must Kantian Constructivists Be?, *Inquiry*, Vol. 49, No. 6.
- Tuck, Richard. *Natural Rights Theories*. Cambridge University Press, 1978.
- Tuck, Richard. *The Rights of War and Peace*. Oxford University Press, 1999.
- Westerman, Pauline C. *Disintegration of Natural Law Theory*. Brill, 1998.
- Wood, Allen. *Kant's Ethical Thought*. Cambridge University Press, 1999.

Notes

- ¹ Kant, G 4:421 and G 4:429
- ² Mackie, Ethics, Inventing Right and Wrong, p 38, quoted in Sources of Normativity, p 37
- ³ I rely on Paul Guyer's account in The Value of Reason and the Value of Freedom in Ethics, Vol. 109, No. 1, 1998
- ⁴ Korsgaard, The Sources of Normativity, p 2
- ⁵ Korsgaard, The Sources of Normativity, p 5. I am not sure that this description of the modern view, that form must be added from without, is entirely adequate, however.
- ⁶ Korsgaard, The Sources of Normativity, p 13
- ⁷ Guyer, *ibid*, p 22-23
- ⁸ Korsgaard, The Sources of Normativity, p 18-19
- ⁹ Korsgaard, The Sources of Normativity, p 90-91
- ¹⁰ Tuck, The Rights of War and Peace, p 34.
- ¹¹ Rousseau, The Social Contract, p 158
- ¹² Quoted in Edmundson, An Introduction to Rights, p 18.
- ¹³ Schneewind, The Invention of Autonomy, p 81.
- ¹⁴ Schneewind, The Invention of Autonomy, p 59.
- ¹⁵ Westerman, The Invention of Autonomy, p 179
- ¹⁶ Tuck, The Rights of War and Peace, p 11
- ¹⁷ Quoted in Schneewind, The Invention of Autonomy, p 68
- ¹⁸ Hobbes, Leviathan, p 1
- ¹⁹ Westerman, Disintegration of Natural Law, p 189
- ²⁰ Westerman, Disintegration of Natural Law, p 197
- ²¹ Westerman, Disintegration of Natural Law, p 224
- ²² Korsgaard, The Sources of Normativity, p 22-24
- ²³ Korsgaard, The Sources of Normativity, p 26-7
- ²⁴ Korsgaard, The Sources of Normativity, p 27
- ²⁵ Korsgaard, The Sources of Normativity, p 29
- ²⁶ Korsgaard, The Sources of Normativity, p 34
- ²⁷ Korsgaard, The Sources of Normativity, p 37
- ²⁸ Korsgaard, The Sources of Normativity, p 40
- ²⁹ Korsgaard, The Sources of Normativity, p 51
- ³⁰ Korsgaard, The Sources of Normativity, p 59
- ³¹ Hume, Enquiry Concerning the Principles of Morals, p 27, quoted in The Sources of Normativity, p 59
- ³² Korsgaard, The Sources of Normativity, p 60
- ³³ Korsgaard, The Sources of Normativity, p 63-4
- ³⁴ Korsgaard, The Sources of Normativity, p 76
- ³⁵ Korsgaard, The Sources of Normativity, p 86-7
- ³⁶ Korsgaard, The Sources of Normativity, p 88
- ³⁷ Korsgaard, The Sources of Normativity, p 89

-
- ³⁸ Korsgaard, *The Sources of Normativity*, p 90-1
- ³⁹ Korsgaard, *The Sources of Normativity*, p 97
- ⁴⁰ Emile, quoted in Schneewind, p 478-9
- ⁴¹ Korsgaard, *The Sources of Normativity*, p 219. She says only that this is Kant's argument, but throughout the text, I will regard her best-effort interpretation of Kant as also being her own opinion, except where she says otherwise.
- ⁴² Korsgaard, *Creating the Kingdom of Ends*, p 23
- ⁴³ Korsgaard, *The Sources of Normativity*, p 98
- ⁴⁴ Korsgaard, *The Sources of Normativity*, p 228
- ⁴⁵ Korsgaard, *The Sources of Normativity*, p 232
- ⁴⁶ Korsgaard, Reply to Carol Voeller and Rachel Cohon, p9. Korsgaard rehashes the argument from Sources here in a slightly different form which she thinks is clearer.
- ⁴⁷ Korsgaard, *Creating the Kingdom of Ends*, p 25
- ⁴⁸ Korsgaard, *Creating the Kingdom of Ends*, p 22
- ⁴⁹ Korsgaard, *Creating the Kingdom of Ends*, p 205
- ⁵⁰ Korsgaard, *Creating the Kingdom of Ends*, p 26
- ⁵¹ Korsgaard, *Creating the Kingdom of Ends*, p 162-3
- ⁵² Korsgaard, *Creating the Kingdom of Ends*, p 183
- ⁵³ Korsgaard, *Creating the Kingdom of Ends*, p 176. My qualification.
- ⁵⁴ Korsgaard, *Creating the Kingdom of Ends*, p 204
- ⁵⁵ Korsgaard, *Creating the Kingdom of Ends*, p 173
- ⁵⁶ Korsgaard, *The Sources of Normativity*, p 98-99
- ⁵⁷ Korsgaard, *The Sources of Normativity*, p 98-100
- ⁵⁸ Korsgaard, *The Sources of Normativity*, p 106-7
- ⁵⁹ Korsgaard, *The Sources of Normativity*, p 101-102
- ⁶⁰ Korsgaard, *The Sources of Normativity*, p 120-1
- ⁶¹ Korsgaard, *The Sources of Normativity*, p 128-9
- ⁶² Korsgaard, *Creating the Kingdom of Ends*, p 16-17
- ⁶³ Korsgaard, *Ethics*, Vol109, p 61
- ⁶⁴ Korsgaard, *Creating the Kingdom of Ends*, p 16-17
- ⁶⁵ Korsgaard, *Ethics*, Vol09, p 64
- ⁶⁶ Korsgaard, *The Sources of Normativity*, p 120-1
- ⁶⁷ Korsgaard, *The Sources of Normativity*, p 123
- ⁶⁸ Korsgaard, *Creating the Kingdom of Ends*, p 123
- ⁶⁹ Korsgaard, *Ethics*, Vol09, p 64
- ⁷⁰ Korsgaard, *Creating the Kingdom of Ends*, p240-1
- ⁷¹ Korsgaard, Interview, p 60
- ⁷² Korsgaard, *Ethics*, Vol09, p 61
- ⁷³ Korsgaard, *The Sources of Normativity*, p 135-6
- ⁷⁴ Korsgaard, *The Sources of Normativity*, p 145
- ⁷⁵ Korsgaard, *Creating the Kingdom of Ends*, p 377

-
- ⁷⁶ Korsgaard, *The Sources of Normativity*, p 153
- ⁷⁷ Kant, MS 8:427
- ⁷⁸ Korsgaard, *Creating the Kingdom of Ends*, p 134
- ⁷⁹ Korsgaard, *Creating the Kingdom of Ends*, p 147-154. See also Interview, p 56-7.
- ⁸⁰ Korsgaard, *Creating the Kingdom of Ends*, p 150-1
- ⁸¹ Kant, G 4:421 and G 4:429.
- ⁸² Korsgaard, *Creating the Kingdom of Ends*, p 153
- ⁸³ Korsgaard, *Creating the Kingdom of Ends*, p 137
- ⁸⁴ Smith, *Internal Reasons*, p 109-112
- ⁸⁵ Korsgaard, *Creating the Kingdom of Ends*, p 151
- ⁸⁶ Korsgaard, *The Sources of Normativity*, p 239
- ⁸⁷ Korsgaard, *Creating the Kingdom of Ends*, p 167
- ⁸⁸ Kant, G 4:450
- ⁸⁹ Kant, MS 6:399, my underscores.
- ⁹⁰ Kant, G 4:460, my underscores.
- ⁹¹ Korsgaard, *Creating the Kingdom of Ends*, p 162
- ⁹² Korsgaard, *Creating the Kingdom of Ends*, p 175
- ⁹³ Tiffany, p 531
- ⁹⁴ Tiffany, p 536
- ⁹⁵ This argument and the regulative versus constitutive distinction is inspired by “How Kantian Must Kantian Constructivists Be?” in *Inquiry* Vol. 49, No. 6., p 524-546, December 2006.
- ⁹⁶ Tiffany, p 525
- ⁹⁷ Kant, G 4:402
- ⁹⁸ Kant, MS 6:399-403
- ⁹⁹ Kant, MS 6:378, MS 6:211-4
- ¹⁰⁰ Korsgaard, *The Sources of Normativity*, p 190
- ¹⁰¹ Korsgaard, *The Sources of Normativity*, p 147
- ¹⁰² Korsgaard, *The Sources of Normativity*, p 239-40
- ¹⁰³ Korsgaard, *The Sources of Normativity*, p 240-1
- ¹⁰⁴ Korsgaard, *Creating the Kingdom of Ends*, p 379
- ¹⁰⁵ Korsgaard, Interview, p 58
- ¹⁰⁶ Hurley, *Consciousness in Action*, p 74
- ¹⁰⁷ Quoted in Hurley, *Consciousness in Action*, p 82
- ¹⁰⁸ Korsgaard, *The Sources of Normativity*, p 241-2
- ¹⁰⁹ Korsgaard, *Creating the Kingdom of Ends*, p 160-2
- ¹¹⁰ Hurley, *Consciousness in Action*, p 80
- ¹¹¹ Korsgaard, *Ethics*, Vol09, p 54
- ¹¹² Korsgaard, *Ethics*, Vol09, p 66
- ¹¹³ Korsgaard, *Ethics*, Vol09, p 51
- ¹¹⁴ Smith, *The Moral Problem*, p 105-108

-
- ¹¹⁵ Smith, *The Moral Problem*, p 112
- ¹¹⁶ Kant MS 6:211-2
- ¹¹⁷ Kant, *Anthropology from a Pragmatic Point of View*, p 155
- ¹¹⁸ Korsgaard, *Creating the Kingdom of Ends*, p 27
- ¹¹⁹ Korsgaard, *Creating the Kingdom of Ends*, p 370
- ¹²⁰ Smith, *The Moral Problem*, p 6 and p 119
- ¹²¹ Smith, *The Moral Problem*, p 7-9
- ¹²² Smith, *The Moral Problem*, p 9-10
- ¹²³ Smith, *The Moral Problem*, p 131-3
- ¹²⁴ Smith, *The Moral Problem*, p 193
- ¹²⁵ Smith, *The Moral Problem*, p 155
- ¹²⁶ Smith, *The Moral Problem*, p 61, and Miller, *An Introduction to Contemporary Metaethics*, p 220
- ¹²⁷ Miller, *An Introduction to Contemporary Metaethics*, p221
- ¹²⁸ Searle, *Rationality in Action*, p 12-13
- ¹²⁹ Smith, *The Moral Problem*, p 106
- ¹³⁰ Searle, *Rationality in Action*, p 111-113
- ¹³¹ Locke, *An Essay Concerning Human Understanding*, book I, chapter II, 7; p 71
- ¹³² Korsgaard, *Creating the Kingdom of Ends*, p 149
- ¹³³ Korsgaard, *Interview*, p 62
- ¹³⁴ Korsgaard, *Creating the Kingdom of Ends*, p 117-8
- ¹³⁵ Korsgaard, *Creating the Kingdom of Ends*, p 169
- ¹³⁶ Korsgaard, *Creating the Kingdom of Ends*, p 275
- ¹³⁷ Korsgaard, *Kingdom*, p 118-119
- ¹³⁸ Korsgaard, *Interview*, p 60
- ¹³⁹ This point is discussed in depth in a debate between Paul Guyer and Christine Korsgaard in *Ethics*, Vol. 109, No. 1.
- ¹⁴⁰ Korsgaard, *Ethics*, Vol09, p 60
- ¹⁴¹ Kant, *Anthropology*, p 175-6
- ¹⁴² Korsgaard, *The Sources of Normativity*, p 127
- ¹⁴³ Korsgaard, *Creating the Kingdom of Ends*, p 166-7. I suspect Korsgaard uses a slightly different definition of self-love here than in later writings.
- ¹⁴⁴ Kant, *Critique of Teleological Judgement*, p 429
- ¹⁴⁵ Kant, *Critique of Teleological Judgement*, p 398-9
- ¹⁴⁶ Kant, *Critique of Teleological Judgement*, p 409-10
- ¹⁴⁷ Kant, *Critique of Teleological Judgement*, p 431
- ¹⁴⁸ Kant, *Critique of Teleological Judgement*, p 441
- ¹⁴⁹ Kant, *Critique of Teleological Judgement*, p 446
- ¹⁵⁰ Kant, *Critique of Teleological Judgement*, p 431
- ¹⁵¹ Kant, *Critique of Teleological Judgement*, p 443
- ¹⁵² Mackie, *Ethics, Inventing Right and Wrong*, p 38, quoted in *Sources of Normativity*, p 37
- ¹⁵³ Korsgaard, *The Sources of Normativity*, p 166

¹⁵⁴ Korsgaard, *The Sources of Normativity*, p 251