

From the Categorical Imperative to the Moral Law

Marianne Vahl

Master Thesis in Philosophy

Supervisor Olav Gjelsvik

Department of Philosophy, Classics, History of Arts and Ideas

UNIVERSITY OF OSLO

May 2008

Synopsis

This thesis is a critical survey of Christine Korsgaard's arguments regarding the rational basis for moral obligations. I focus on her arguments taking us from the categorical imperative to the Moral Law. She makes a distinction that Kant does not; claiming that the categorical imperative is not the Moral Law. In order to equate rational agency with moral agency Korsgaard therefore needs some additional arguments. These arguments, I argue, are not convincing. My claim is that they do not succeed in establishing the necessity which Korsgaard actually attributes to moral obligations, nor support her idea that our moral identity is inescapable.

In the first chapter, I give an interpretation of her view pointing to some similarities and differences between her arguments and those of Hume and Kant. Central to my discussion is the tension in her position due to arguing in accordance with Hume that morality is grounded in human nature, and at the same time arguing in accordance with the Kantian idea of autonomy as the answer to our quest for responsibility. Her view is vague concerning the important distinction between rational nature and human nature; a vagueness that is accentuated by her introduction of the notion of procedural realism.

Both Kant and Korsgaard are commonly taken to be constructivists. In the second chapter, I explore this aspect of her view by taking into account some constructivist ideas as presented by Scanlon and Rawls. I argue that Korsgaard, unlike Scanlon and Rawls, takes constructivism too far. She distinguishes between our third personal perspective of explanation and our first person perspective of deliberation: It is from the first person perspective of deliberation that we both justify and construct central and related concepts like the categorical imperative, morality, our identity, normativity and value. It is difficult to see how she can argue in this manner without running into problems concerning self-reference and circularity.

A key problem is her conception of deliberation as a procedure separating our inclinations from our reasoning. In the third chapter I therefore explore further her conception of deliberation, and in particular the role emotions may play. I also compare her notion of identity with Allison and Scanlon's conception of a self, in order to support my suspicion that her view is based on some sort of naturalism. I argue that the duality she requires from our deliberate standpoint puts an unnecessary strong demand on the causality of human reasoning. But is naturalism the only plausible explanation of how Korsgaard's claim for necessity can be met? Perhaps she does what Kant did; justify necessity by transcendental arguments?

In the last chapter, I explore whether or not her arguments are transcendental, and I conclude that her arguments cannot establish the necessity she requires from our obligations. By distinguishing between the categorical imperative and the Moral Law, she opens up for the possibility of autonomous agency equating rational agency, rather than moral agency as she had set forth to show. It is actually questionable whether she manages to establish that rational agency is autonomous too. But why should the fact that we are the source of the law that binds us, necessarily lead to freedom and responsibility? For Kant and Korsgaard this seems to be the solution to the problems arising out of seeing nature as deterministic and having an absolutistic conception of egoism. If one did let go of these – in my opinion unnecessary – worries, autonomy is perhaps not the answer to our quest for responsibility.

Acknowledgements

First and foremost I would like to thank my supervisor Olav Gjelsvik for all his valuable feedback, and for seeing the potential in my writings at times when this was far from obvious. I am also very grateful that Heine A. Holmen took time to read and comment on an earlier version of this thesis. Further, I want to thank my friends and family for all of their support. In particular, I am grateful to Gabriel for all support and inspiration. But most of all I am forever grateful to my children Filip and Rebekka for their understanding, patience and love which encouraged me to complete this thesis.

Table of Contents

	Page
1.0 From the categorical imperative to the Moral Law	5
1.1 The quest for responsibility	5
1.2 The categorical imperative as the law of a free will	7
1.3 Kant's different formulas of the categorical imperative	8
1.4 Reflective endorsement and practical identity	13
1.5 Seeing the value of one's own humanity	17
1.6 The publicity of reasons and the value of others	18
1.7 Double- versus single-level theories	23
1.8 Conclusion	24
2.0 Justification and explanation	26
2.1 Coherence theories	26
2.2 Procedural realism	28
2.3 Original position argumentation	29
2.4 Justification of the categorical imperative	30
2.5 Reflective endorsement – the procedure of the categorical imperative	33
2.6 Personal identity and unity of agency	36
2.7 Two standpoints and the Modern Scientific World View	38
2.8 Pre-reflective intuitions and the unconditional	41
2.9 Conclusion	45
3.0 Deliberation and the dualities of reason	47
3.1 Korsgaard's conception of rationality	47
3.2 Scepticism about practical reason	49
3.3 Reasoning and emotions	50
3.4 Naturalism and the nature of the self	57
3.5 Conclusion	61
4.0 Autonomy and freedom	62
4.1 Transcendental argumentation	62
4.2 Transcendental argument for the value of humanity	65
4.3 Universalizability and the necessity to see the value in others	67
4.4 Spontaneity, autonomy and authorization	72
4.5 The value of choice and the causal thesis	74
4.6 Conclusion	76

1.0 From the categorical imperative to the Moral Law

In this thesis I intend to present a critical survey of Christine Korsgaard's arguments taking us from the categorical imperative to the Moral Law. The law which can be understood as telling us that we must take other peoples reasons into consideration in situations of choice and action. I will primarily base my interpretation on her writings in *Creating the Kingdom of Ends*¹ and in *the Sources of Normativity*². In these books Korsgaard can be understood as arguing for a rational basis for morality in terms of obligations. To establish the Moral Law, she presents arguments that may be said to go through three phases³: the foundation of the categorical imperative, to see the value in one's own humanity and to recognize the value in others. Korsgaard's arguments may be seen as a response to Kant, claiming what he did not; that the categorical imperative is different from the Moral Law.

In this first chapter my aim is to present Korsgaard's discussion of Kant's view with regards to the relation between the categorical imperative and the Moral Law. Then I will critically explore her view on the very same issue. In the following chapters I will discuss aspects of her arguments which I find problematic or unclear; difficulties mainly due to the claim of necessity Kant attributes to both the categorical imperative and to moral obligations. This claim, I will argue, Korsgaard's view does not obviously meet.

1.1 The quest for responsibility

Despite their differences, I find that both Kant and Korsgaard appear to take as a starting point the quest for responsibility. To consider human beings as not responsible for their thoughts and actions may seem like both an unappealing and impossible idea if at the same time, we are to view ourselves as having some kind of identity that has an impact on our lives. However, to argue for a view that holds that we actually are responsible also gives rise to several challenges, challenges often the result of the commonly assumed connection between responsibility and the idea of a free will. In *Groundwork of the Metaphysics of Morals*⁴, Kant poses the following problem:

¹ Christine Korsgaard, 1996b, *Creating the Kingdom of Ends*.

² Christine Korsgaard, 1996a, *The Sources of Normativity*

³ Stern does in "Transcendental argumentation" argue for this division of her argument

⁴ Kant, *Groundwork of the Metaphysics of Morals,...*, trans. Mary Gregor, Cambridge University Press 1997.

“there arises a dialectic of reason since, with respect to the will, the freedom ascribed to it seems to be in contradiction with natural necessity” (Kant 1997, p.60, 4:455)

Kant seems to both hold that we are rational beings with a will that *acts under the idea of freedom*⁵ and, at the same time, when reasoning about the possibility of experience we realize that everything is determined by the laws of nature. Thus our reason appears to give rise to the contradictory beliefs that we, on the one hand, have a will and in that sense our actions are free and undetermined while, on the other hand, as agents in the world of nature our will is also causality and must as such have some law or principle in accordance with the laws of nature. However, what is the source for such a law if it is not to be determined by external powers, and how is it possible for freedom to imply law? Kant’s answer to this is that there still is the possibility that we can make laws to ourselves – that we are autonomous⁶. Hence the apparent contradiction is to him not a real problem. Our will is autonomous and its law is what Kant called a categorical imperative, formulated by the 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.”(Kant 1997, p.31, 4:421)

Kant is commonly understood to equate this formula with what he refers to as the Moral Law, this being due to the idea that universal ranges over all rational beings. To understand Kant’s arguments is a complex task that has given rise to a variety of interpretations. I will not try to justify these here, but rather focus on the view of Christine Korsgaard who poses a similar problem to Kant. However, her solution differs, or as she says⁷ ‘universalizability does not get us to morality’ since universalizability cannot bridge the gap between what is a reason for you to act from, from what is a reason for me to act. Thus Korsgaard wishes to improve on this often considered weakness in Kant’s argument by presenting some additional arguments to the Kantian ones in order to bridge this gap from the free will to moral obligations. Here, I will critically explore these arguments given by Korsgaard and first begin by looking at the area in her argumentations where she claims to agree with Kant – namely

⁵ Kant, *Groundwork of the Metaphysics of Morals*, p.53, trans.Mary Gregor, Cambridge University Press 1997.

⁶ Autonomy as the solution to this problem will be critically discussed in the last chapter.

⁷ Korsgaard 1996a, p. 221.

on how to establish that we are autonomous beings. Second, I will explore her argument from the categorical imperative to the Moral Law; an argument which seems to rest upon a un-Kantian assumption that the content of the Moral Law is based upon what she calls our conception of our practical identity. Only by identifying yourselves normatively with others, by having the self-conception of being a member of a social community, can the law given by the categorical imperative be the Moral Law she claims.

1.2 The categorical imperative as the law of a free will

A major concern in most ethical theories is how to deal with the concepts of the good and the right, and in particular in what way these concepts are related. Since ‘the good’ may say to be about how we want things to be and “the right” on what we should or must do, both concepts are important due to their alleged motivational force on our decisions and actions. Teleology⁸ is the view often associated with attributing values to state of affairs, or on how things may become. These are matters with their own independent value. Often is the teleological view combined with a consequentialist view, a view that emphasizes the importance of the consequences of an action, consequences which may be stated as facts. Korsgaard disagrees with the teleological conception – at least as telling the whole truth. Instead Korsgaard argues in agreement with the following interpretation of Kant: In contrast to Aristotle and Plato, Kant said, that “reason - which is form - isn’t in the world, but is something that we impose upon it.”(Korsgaard 1996a, pp.4-5). Thus if the good or form is not considered what is real, where does this leave the issue of value? According to this view, value is imposed upon the world of matter; it is an obligation. Thus rather than searching for value, we create value in the world around us, forced by our obligations - it is like a work of art - Korsgaard explains. As rational beings, we determine our ends by being the ones who determine value, and we have duties to have these ends. According to her, the underlying idea is that “the reason why a good-willed person does an action, and the reason why the action is right, is the same” (Korsgaard 1996b, p. 61). A good-willed person is not motivated by private purposes, but acts from necessity that it is a law to have a certain purpose. Korsgaard says that Kant distinguishes between two kinds of

⁸ Including the view of Aristotle.

motivation; autonomous and heteronomous. To be heteronomously motivated is to act according to some interest, where a hypothesis connects your interest to the law which binds you. The command to perform the action is hypothetical in the sense that we can choose to give up the action by changing our interests. However, Korsgaard continues, we need to be unconditionally bound by an imperative - an ought - on what to do otherwise duty does not obligate us by necessity. Moral motivation has to be autonomous, that is to be autonomously motivated is understood as acting on laws you have set to yourself. Which acts are right depends on our maxims, where maxims may be understood as our underlying aims, intentions, subjective principles or reasons for acting; it is our inclinations such as those of desires, tastes and feelings. However a rational or good-willed person should not act upon a maxim unless it can also serve as an imperative categorically, meaning by necessity and independently of private concerns. This is the idea of the *categorical imperative*, the principle of autonomy, as we recall as formulated in Kant's 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.” (Kant 1997, p.31, 4:421)

The categorical imperative is the law of the free will Korsgaard continues, because all it says is that we should choose a law. Thus it places no alien constraints to the free will, merely says what the will must do in order to be a free will – it must will a maxim to be a universal law. The categorical imperative is not an analytic (by definition) judgement, but synthetic a priori in the sense of telling us something substantial, something new about its subject, but still based solely upon reasoning and not upon experience.

1.3 Kant's different formulas of the categorical imperative

How to link the categorical imperative to the Moral Law is an issue where Korsgaard's view differs radically from Kant. I will now present Korsgaard's critical interpretation of Kant on this as given in *The Sources of Normativity* and in *Creating the Kingdom of Ends*.

According to Korsgaard, the concept of freedom is introduced by Kant as a solution to a problem⁹; the categorical imperative is not analytic and as such not inconsistent whatever claims are made, and at the same time it is claimed to consist in rational necessity Korsgaard explains. To answer this one must first show that a free person is one who follows the Moral Law then show that a rational person considers themselves as being free. To the first question concerning how we can come from the Kantian argument from free will to morality Korsgaard finds that Kant has a good answer. More problematic she - and according to her also Kant, and several of his interpreters - finds the relation between freedom and rationality and the argument that being rational implies acting in accordance with the Moral Law. To explain this Kant puts forward two very different arguments¹⁰, one in the *Groundwork* which he denotes a deduction, and the other in the *Critique of Practical Reason*. Korsgaard¹¹ reconstructs Kant's argument given in the *Groundwork* in short as follows: the existence of a categorical imperative implies that there must be something unconditionally valuable, something being an end to itself and as such could be the source of laws. This end must be completely justified and be a necessary end to every rational or good will. According to Korsgaard Kant equates humanity with such a rational will, so what is unconditionally valuable is humanity. Hence the unconditional end required for the categorical imperative to determine the will, is humanity. This, Korsgaard claims, brings Kant to another formulation of the categorical imperative, the Formula of Humanity:

“So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.” (Kant 1997, p. 38, 4:429)

This is a very strong requirement - to never *merely* treat someone as means. This requirement appears to be based on a view that egoism in an absolute sense is the challenge. This in my opinion is an unnecessary strong demand to put on our actions. I will return to this discussion later. First I will continue my exploration of Kant's arguments: So if rational beings determine ends, and are the sources of value, rational

⁹ Korsgaard 1996b, p. 160.

¹⁰ Allison, *Kant's theory of freedom*, p. 2, Korsgaard 1996 b), p. 161.

¹¹ Korsgaard, 1996b, pp.12 – 27.

beings must value themselves and as such consider themselves as ends. Rational beings are autonomous, and rational nature is an objective end. We consider other humans as important by worshipping ourselves and our humanity, which is what may move us to act morally. It is about respect for human beings, but only insofar as they are rational, and we will be moved only insofar that we are rational. So what ultimately motivates us is the emotion respect, respect for the moral law and for ourselves and other rational beings. This leads Kant, according to Korsgaard, to the idea of an ideal community where everyone considers one another as ends in themselves – a community he called *The Kingdom of Ends*. With this he presents the third alternative way of formulating the categorical imperative:

“A rational being must always regard himself as lawgiving in a kingdom of ends.”(Kant 1997, p. 42, 4:434)

Hence the law which we would rationally choose when being governed by the categorical imperative is the Moral Law. By stating this Korsgaard argues that to Kant the Formula of Universal Law is the Moral Law. It is a practical law applying to all members of the kingdom of ends, and to follow this principle is *practical necessity* – it is a duty. But how can universalizability lead us to the Moral Law, Korsgaard asks? Why does it have to range over human beings? Does the Formula of Universal Law have moral content, or is it perhaps what several critics suggest, an empty formalism¹²? If this empty formalism is equated with the Moral Law, does it leave the Moral Law empty as well? Perhaps by investigating how one is supposed to apply the categorical imperative one can answer these questions.

Korsgaard argues that the claim that the moral law is empty is wrong. According to her Kant argued that if all rational beings could agree, that such a maxim could be a law for all members of the Kingdom of Ends, then the categorical imperative would have content. But how can you tell whether you are able to will your maxim as a universal law, Korsgaard asks¹³. Kant’s answer to this question is, according to Korsgaard, to seek universalization without contradiction. By imagining a world in which a maxim is being universalized, Korsgaard interprets Kant as meaning a world in

¹² ‘No content’ objections have been made by Mill, Hegel and his followers. (Korsgaard 1996a, p. 222).

¹³ Korsgaard, *Creating the Kingdom of Ends* p.14.

which the maxim is envisioned as a law of nature. The arguments which Kant gives to show that our maxims must undergo a test of universalizability are summarized by Korsgaard in the following way¹⁴: we must act under the idea of freedom or under the idea that we have a free will. This means that we act as if we were not determined by any external (to the will) forces or laws. However, Kant defines a free will as rational causality, meaning something which is a cause without being determined by anything alien to it. As alien to our will all our inclinations and emotions are included as well as issues completely external to us. This leaves us with a problem: how can the will be completely self-determining and, at the same time, be causality and as such act in accordance with laws? Kant's answer to this is that the will is autonomous – it has its own law. This gives rise to a second problem. Where does this law come from? It cannot be external to the will, since then the will is not free. Thus the will must make a law to itself, and the law of the free will is the categorical imperative (as the law of autonomy) as given in the Formula of Universal Law. Hence he proposed a test seeking to establish what one can *will without contradiction*.

“To determine whether you can will your maxim at the same time as its universalization without contradiction, you envision trying to will your maxim in a world in which the maxim is universalized – in which it is a law of nature.” (Korsgaard 1996b, p.14)

In what she calls “the Practical Contradiction Interpretation” our will is causality, and we must consider the relation between our reasons or purpose to perform an action and the action as universalizable, as a law, without contradictions, and if not our reasons are not sufficient¹⁵. Our non-moral maxims cannot in this view serve as practical laws. Thus Korsgaard argues that the contradiction test manages to provide some of the content of morality, since it is capable of deciding the moral content of some maxims. Hence it shows that the Moral Law is not empty. But, she continues, it does not manage to give us the whole content of morality, and more importantly it does not establish that being bound by the categorical imperative implies that you are bound by the Moral Law. The Kantian argument shows the categorical imperative to be the law of a free will, but it does not establish the same for the Moral Law¹⁶.

¹⁴ Korsgaard, 1996a, p.219, p97.

¹⁵ Korsgaard 1996b, p.102.

¹⁶ Korsgaard 1996a, p. 99.

Thus in this respect Korsgaard's view diverges from Kant. The difference is due to the domain of the law constituted by the free will. According to Korsgaard, the law of the free will does not necessarily include the Kingdom of Ends: it is not decided whether the categorical imperative range over rational beings, human beings, moral agents or something else. According to Korsgaard, the categorical imperative is the law saying that you should only act on maxims you would want to be a law. It is a law of the free will in the sense that all it asks from us is to choose a law, but it places no constraints on how this law should be; rather, it puts constraints on its form. The Moral Law, on the other hand, does place such constraints on the content of the law; it demands that one must agree to the law. But does not the universalizability requirement in the formula of universal law imply that the law is for all rational beings? No, Korsgaard replies¹⁷. All that universalizability requirements do is to say that if something is rational for me to do for egoistic purposes, then I must agree that it would be rational for you to go for your interests. However, this does not make your interest normative to me. What does this really mean? That by universal Korsgaard means general, in the sense of something which will always have a hold on me, as a law in the sense of 'always' and not in the sense of 'for everyone'? She appears to consider universal as something which could apply to only one person, but which probably would also be attractive for others in the same situation. Does this mean that the categorical imperative could have a hold on us, without having other human beings within its domain and, if so, would this action be rational? I believe she has to answer yes to the first questions; there will be occasions where something is a reason for me without being a reason for others. Korsgaard's distinction between the categorical imperative and the Moral Law seems to make some acts of egoism rational. But, if it seems rational for me to follow a principle, it will probably seem rational for another rational being as well. However again, this does not, according to Korsgaard, give us morality¹⁸. This makes one suspect that she may hold another conception of rationality than Kant, a conception seeing rational and universal as distinct matters. One way to avoid this problem concerning how to come from egoism to morality, is of course, to interpret universal differently than Korsgaard. This is exactly what Tim Scanlon is interpreted as doing in *What We Owe to Each Other*. He argues that a universal law is a

¹⁷ Korsgaard 1996a, p. 221.

¹⁸ Korsgaard 1996a, p. 221.

law for everybody¹⁹ and not as Korsgaard seems to mean general in the sense of always. However, his view is heteronomously based, and as such I guess Korsgaard would argue cannot provide the necessity she requires from moral obligations. Her claim that all the categorical imperative is, is a law, which has no content other than structure. Korsgaard concludes that universalizability does not imply morality. But how then does she argue her way from the categorical imperative to the Moral Law? To answer this she presents a two-fold argument:

“first, that the universal laws required by our conception of ourselves as agents must range over human beings as such; and second that the reasons that are derived from these laws are public.”(Korsgaard 1996a, p. 233)

Hence, in order to come to morality one needs a further two arguments in addition to the categorical imperative Korsgaard claims. First, she wishes to establish that rational actions exist and are related to placing value in one’s own humanity. Second, she gives an argument equating valuing one’s own humanity with seeing the value in the humanity of others. To give an account of Korsgaard’s arguments from the categorical imperative to the Moral Law requires that I say something about reflective endorsement and practical identity, so I start by doing that and will then outline the arguments.

1.4 Reflective endorsement and practical identity

The Kantian test as to whether or not we can act on some maxim, depends on which maxims we could will as a universal law. The procedure he uses according to Korsgaard, is one of reflective scrutiny, aiming at rejecting maxims which lead to some sort of contradiction. This procedure, she continues, is similar to the procedure of ‘reflective endorsement’²⁰, a method she attributes to Hume, Mill and Bernard Williams. Korsgaard explains that for Hume because that morality has its foundation in human sentiments, the normative question is how good morality is and not whether it is true or not. Bernard Williams, she continues, is likewise concerned about whether we have reasons to endorse our dispositions. Even Mill uses the method of reflective endorsement when answering the question of normativity of our obligations. Thus,

¹⁹ Parfit What We Could Rationally Will.

²⁰ Korsgaard 1996a, pp.49-130.

according to Korsgaard, the reflective endorsement method is one of reflection, based upon the idea that the ultimate source of which reasons we have for action lies in human nature. But what does she mean by that? To claim that Hume's theory is normative, Korsgaard admits, is a controversial claim. She bases her view on the fact that Hume's division between the ways we deal with moral philosophy into theoretical and practical philosophy. From a theoretical point of view what we want is to explain moral concepts, while the practical philosopher wishes to persuade and convince people to behave in certain ways. But then, how does one deal with justification? For this Korsgaard attributes to Hume the idea that it will emerge by the interaction between the theoretical and practical views of philosophy. So in this sense Hume's theory is normative according to Korsgaard. Human nature is *intrinsically normative*, so morality can only be challenged from the standpoint internal to human nature. From this internal standpoint morality and self-interest are coherent notions, since we have (or perhaps couldn't have?) no reasons to disapprove of morality since it is our nature. What Korsgaard seeks is a test for normativity in order to decide what we can take to be a reason for action. When through reflection we reach endorsement concerning our reasons, we approve or adopt these reasons. Korsgaard's project is to argue that "the logical consequence of the theory of normativity shared by Hume, Mill and Williams is the moral philosophy of Kant"²¹. But in what sense can the view of Hume lead to Kant's view? By claiming this I believe that one problem is that the method of reflective endorsement seems to be based on the idea that the source of normativity is in human nature, and not in rational nature which is how Kant is commonly understood²².

"Kant, like the realist, thinks 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. But the test that it must pass is not the test of knowledge or truth. For Kant, like Hume and Williams, thinks that morality is grounded in human nature, and that moral properties are projections of human dispositions. So the test is one of reflective endorsement."(Korsgaard 1996a, p. 91)

Here Korsgaard comes with the somewhat controversial claim that Kant grounds morality in human nature. Does this mean that Korsgaard equates human nature with

²¹ Korsgaard 1996a, p. 51.

²² Cohen gives a similar reply to Korsgaard in *Reason, Humanity, and the Moral Law*, Korsgaard 1996a pp.167-188.

rational nature, and that she also believes that Kant did? If this is the case, what does she really mean by conceptions like autonomy, the good will, rational will and freedom? Not obviously the same I believe Kant had in mind, as something based upon what transcend human nature.

Anyway, both Kant and Korsgaard seem to argue against the view that morality is based upon truths and objective values, and that knowledge is the source of normativity. Rather Korsgaard claims, the source of normativity is in human nature, and so what we need for normativity is confidence – not knowledge. Korsgaard argues that one of the things which distinguish us from other living creatures is exactly that we are capable of being self-conscious. Our mind is reflective in the sense that we can both pay attention to our mental activities and observe them from a distance. When we find ourselves with an impulse to act, such as a desire or a perception, we step back and evaluate whether these impulses really are reasons to act or not. The reflective mind requires reasons in order to act, and to be aware of such a reason means one has been successful in reflection. This reflective structure of our mind and, our self-conception possible because of it is what Korsgaard considers to be as being the conception of our own practical identity; practical because of our ability to identify our reasons and to identify within the source of those reasons, reasons about what to do. So, it is not what would have been the aim of theoretical reflection to find out what is in the normative part of the world. It is this conception of us, as a practical identity, that gives rise to obligation. Obligations are a reaction to what your practical identity will not allow to happen. If you violate your self conception you loose your identity, so obligations arises when your identity is threatened.

“If reasons arise from reflective endorsement, then obligation arises from reflective rejection.”(Korsgaard 1996a, p. 102)

The reflective structure of our consciousness implies that there must be some laws which make us command ourselves, which makes us obliged. Korsgaard continues to argue that the source of obligation is our autonomy, since it demands that you are a law unto yourself. Thus human nature - the reflective structure of our consciousness - is the source of normativity. However one of the main problems we may seem to have with regards to normativity is precisely due to this practical reflective nature. When we are

inclined to believe that that we ought to do something we can always question this conviction, we can wish to justify morality's claims on us. She continues that if morality survives this kind of scrutiny, then morality is normative. This is the method she refers to as the 'reflective endorsement method'.

Where Hume and Williams seek to establish the normativity of our "moral dispositions and sentiment"²³ Korsgaard considers Kant as going further than merely seeing it as a test justifying morality - *it is morality*. Hence Korsgaard continues, that what seemed to be a problem to normativity, our reflective nature, turned out to also be the solution to the problem. Morality is a process, a reflective one. The reflection in question according to Korsgaard is practical and concerns what to do. It is not theoretical, in the sense of reasoning about normativity. We distance ourselves from our motives, perceptions and desires – all of our inclinations - and by so doing, are able to choose between them. Thus being self-conscious in this sense, capable of reflecting about our own mental activities, is, she claims, the very expression of your will. It shows that you are not passively letting your desires decide what you do, instead you are actively choosing. This she says is because you have reasons for acting. However, that we are bound by the categorical imperative does not establish that we are bound by the moral law according to Korsgaard. To establish this we need to show that we have a conception of our self as a member of the Kingdom of Ends²⁴. Korsgaard claims Kant does this by saying that our mind has a reflective structure which is the source of our self-consciousness. According to Korsgaard, Kant does not find this an argument in favour of the existence of a metaphysical self but rather as concerning what the deliberate process is like from a first person perspective.

Korsgaard draws this line of thought further by claiming that when we deliberate, it is as if there is something disconnected from our desires, some kind of ego making our choices. According to Korsgaard human beings have a reflective nature, but to reflect requires some conception of a self, an identity. It is this identity that provides us with the reasons we need in order to act. What we take to be a reason depends on our practical identity, and we have several such identities she continues. For example, we can have the practical identity of being a mother, a teacher, an European, and all other different roles we may have. So, what count as a reason depends on these practical

²³ Korsgaard 1996a, p. 89.

²⁴ Korsgaard 1996a, p. 100.

identities, in this sense what is a reason for me is not a reason for you. Our moral identity is the identity where we identify ourselves with others in the Kingdom of Ends. This moral identity is hence one of our practical identities, but different from our other identities in that it is inescapable Korsgaard claims. Any self conceptions inconsistent with you seeing yourself as a member of humanity must be avoided. Korsgaard says the feature most central to this necessary identity is that of being able to reflect – to be self-conscious. Thus to avoid the scepticism of practical reasons humanity must be valued as such²⁵. In this sense Korsgaard appears to equate moral identity with valuing one's humanity, but is this not the same as claiming that our rational reasons are the same as our moral reasons? And how can she claim this, and at the same time argue for her understanding of universal as not necessarily ranging over other people. This, as I have argued, opens up for having reasons (rational) which are not moral reasons. How does she argue in order to show that our rational reasons are in fact our moral reasons? How does she justify that to see the value in ones own humanity entails seeing the value of the humanity in others?

1.5 Seeing the value of one's own humanity

In summary, the arguments for seeing the value of one's own humanity may be said to be something like as follows²⁶: Action requires that we take something to be a reason to act. This means that I identify with my principle of choice; I have to see myself as the agent of the action. What I take as a reason must conform to my practical identity. Humanity is a practical identity to which all other identities must conform. To have this self-conception is due to valuing your reflective powers, which means valuing your humanity. You have to value your humanity, Korsgaard concludes.

Thus Korsgaard claims that we must necessarily have a conception of a practical identity, our identity of humanity, in order to have reasons to act at all. Owing to being reflective agents capable of being self-conscious, we “are forced” to have a conception of ourselves. The source of normativity is reflection - it is human nature, she continues. It is by valuing your own humanity contingent values becomes necessary values. Hence

²⁵ Korsgaard 1996a, p.122.

²⁶ A similar presentation is given by Stern in "The Value of Humanity: Reflections on Korsgaard's Transcendental Argument

Korsgaard seems to ground necessity in the value of human nature, to the value of our reflective and rational powers. I suppose that it is in this sense that she equates human nature with rational nature, that the essence of both human and rational nature is reflection and the ability to be self-conscious. Thus, to consider human nature and rational nature as somewhat the same appears to be a necessary condition for this argument. If this is right it identifies a central difference between Kant and Korsgaard, associating Korsgaard's view more with the view of Hume than that of Kant. Furthermore, this argument does not rule out egoism. An egoist could both accept instrumental reasoning as means to an end and value his own humanity. Further, an egoist may acknowledge the same features in others, that is acknowledge the egoism of others. What the egoist does not is to, by necessity value other people's humanity and promotes their ends²⁷. Korsgaard's reply to these concerns is that the reasons of the egoist are private – or agent-relative - and to count as reason it must be public. So what does that mean, that Korsgaard thinks that egoism is impossible? Does valuing your humanity imply that you value the humanity in others? How do we come from our valuing our humanity, to valuing the humanity in others? To accomplish this Korsgaard one more argument is needed.

1.6 The publicity of reasons and the value of others

“The space of linguistic consciousness – the space in which meanings and reasons exists – is a space that we occupy together.”(Korsgaard 1996a, p. 145)

We are obligated by reflection, Korsgaard says. Thus if someone is going to obligate me, it will require that I become aware of them, that they could somehow be involved in my reflections. We need to show that our reflections are not private – that reasons are not only normative for me. Korsgaard parallels her argument with that of Wittgenstein's private language argument showing the normativity of linguistic meaning. The idea is that when our reflections result in formulations of which reasons we have (maxims), these are formulations that require language. Our maxim is communicable and, hence,

²⁷ James Skidmore gives a similar explanation of an egoist in his paper “Scepticism About Practical Reason: Transcendental Arguments and their Limits”, p. 135.

not private in the sense that they are intelligible to others. Thus to value one's own humanity or agency is communicable, and intelligible to others, and hence implies the value of humanity in general. This Korsgaard concludes, shows that we have moral reasons – owing to reasons being public. This ends her argument which took us from the categorical imperative to the Moral Law. However this argument is much criticized and among the problems discussed is how to understand Korsgaard's conceptions of private and public. Are public reasons to be understood as reasons which we actually share or are they only to be considered as shareable? Can norms be private, can the egoist have obligations towards himself? Theo van Willigenburg²⁸ presents in the paper "Shareability and Actual Sharing: Korsgaard's Position on the Publicity of Reasons" an interpretation of Korsgaard as different from Kant due to Kant meaning that "reasons are necessarily shared because they cannot but count for every citizen of the Kingdom of Ends" (Willigenburg 2002, p.173). Korsgaard on the other hand claims according to Willigenburg that there is a difference between understanding publicity as meaning actual sharing or as meaning intelligibility. It is in making this distinction Korsgaard can claim that the argument for the categorical imperative does not take us to the Moral Law. I find this a plausible reading of Korsgaard, and will explore it further. It is a reading not making the Kantian assumption that shareability of reasons entails actual sharing. Or, at least to establish this relation one needs a further argument, Korsgaard claims.

"It is because the standpoint created by consciousness can be made public by language or sympathy that reasons and values can be shared. But that kind of publicity is still inside the reflective standpoint. From outside of that standpoint, we can recognize the fact of value, but we cannot recognize value itself."(Korsgaard 1996a, p. 161)

Is this to be understood as some sort of internal publicity? In that case what she seems to argue for is the ability to share and not actual sharing as sufficient for also valuing the humanity in others. But is this really enough for us to be morally obligated, that we, by reflection, can understand that something is a reason for others? I interpret Korsgaard as meaning exactly this – reflection is the source of obligation: a view I assume is based on the idea that we have some common sense concerning what is intelligible.

²⁸ Theo van Willigenburg, "Shareability and Actual sharing: Korsgaard's position on the Publicity of Reasons", pp.172-174.

“The private language argument does not show that I could not have my own personal language. But it shows that I could not have a language that is in principle incommunicable to anybody else. When I make a language, I make its meaning normative for me.” (Korsgaard 1996a, p.138)

But, to make a language meaning normative to me does not in my opinion imply that this meaning needs to have to do with others, its meaning could very well be related to my own egoistic concerns. I will therefore argue against that the argument given by Korsgaard establishes what Willigenburg claims that it does, namely that “we cannot but share what is already inherently shareable”²⁹. My argument will to some extent be in accordance with the ones of James Skidmore. In his paper “Scepticism about Practical Reason: Transcendental Arguments and their Limits” he argues³⁰ that with an incommunicable language, I could not make mistakes, so there would be no normativity. But why should we only be concerned about moral ends? What about the rational ends - are they necessarily the same as our moral ends? With an incommunicable language I find that the problem concerning rational ends would rather be that we could end up always making mistakes, which, of course, would also be a problem to normativity – but of a very different kind. We would end up misunderstanding our own reflections. Korsgaard apparently argues in accordance with Skidmore on this, claiming that egoist reasons are agent-relative reasons, and cannot be normative. In order to be normative reasons they must be agent-neutral, or as she says – public. But why is this so? I do not think this rule out the normativity of egoists’ reasons, since an egoist could have a communicable language but choose not to communicate. Korsgaard would argue against this, saying that one can always intrude into someone’s consciousness, simply by talking to them. In this sense our consciousness is not private. However, I will argue, this is not always the case, and not for all of one’s reasons. Thus in my opinion, private reasons are still possible, and may be normative. This argument does not take Korsgaard from shareability to actual sharing for all of my reasons, only perhaps for some. I find that one of the problems with this argument is due to her conception of egoist, of which she says that: “if egoism is true, and reasons cannot be shared” (Korsgaard 1996a, p. 141) Which I find to be a

²⁹ Willigenburg 2002, p. 189.

³⁰ Skidmore 2002, pp. 135-137.

too strong demand of egoism for an egoist, that one either are an egoist or one are not, that ones an egoist always an egoist. The claim made by the Formula of Humanity that one must never merely treat others as means is too demanding. What about the more common situation in which sometimes one behaves egoistically and other times not? In this reading of Korsgaard's conception of egoism, it seems to be like an inescapable identity - an identity contrasting the moral identity.

Hence, I find that egoism is not ruled out by her argument, but then I mean egoism as understood in a weaker sense than what Korsgaard appears to mean. This I find to coincide well with her claim that the categorical imperative is different from the Moral Law. This division opens up the possibility that the categorical imperative may have a hold on us, without involving any moral claims. Thus, normativity in this necessary sense does not need to be moral, but can be solely rational. This again, I assume, means that one can imagine the rational egoist who can do wrong, not towards others, but towards his own rational ends. Normativity is thus possible due to the argument that one values one's own humanity. Hence, I believe that what we have here that shareable is not the same as actual sharing in the sense that my reasons are always reasons for others. Morality is not intrinsic to my reasons. But if we did accept Korsgaard's claim that to value your own humanity is to value the humanity in others, how does Korsgaard come from this to moral obligations? To do this she claims to follow an argument of Thomas Nagel³¹:

“You make yourself an end for others; you make yourself a law to them. But if you are a law to others in so far as you are just human, just someone, then the humanity of others is also a law to you. By making you think these thoughts, I force you to acknowledge the value of my humanity, and I obligate you to act in a way that respects it.” (Korsgaard 1996a, p. 143)

Hence obligations are constructed by reflection, reflection which somehow forces us to take others into account. By identifying yourself with others, seeing what everyone has in common makes one take others reasons into account. It is by adopting the perspective of others that you become obligated, by reflecting and acknowledging the value of someone else's humanity she continues. Imagine that we are reasoning together, sharing decisions by joining wills, Korsgaard says. Why, Korsgaard asks, should it be

³¹ Korsgaard 1996a, p.142, refers to Nagel's argument in *The Possibility of Altruism*.

impossible to think that language can force us to union in both our thinking and in our practical reasoning? But how can she claim that this happens by necessity? Once again, I believe this is an argument of the shareability of our reasons and not necessarily what actually become our reasons for acting. I believe that the moral identity is escapable. This is in contrast to what Korsgaard means when claiming that “To treat your human identity as normative, as a source of reasons and obligations, is to have what I have been calling ‘moral identity’” (Korsgaard 1996a, p.129) where it seems that reasons and obligations are intimately connected if we are to consider our identity as normative. That we must have such a moral identity does not mean that it is the only identity; no one is only a moral agent, Korsgaard says. Moral obligations are not our only obligations, and do not always win³². Its superiority lies in that it cannot be denied unless one also rejects normativity and practical reasons. So moral obligations don’t always win, meaning egoism is possible. However, egoism is not normative; since egoist reasons are not reasons as such is what Korsgaard seems to believe. And, more importantly: that your moral obligations do not win; do not, as I read Korsgaard, imply that you lose your moral identity. I believe what she has in mind is rather to allow for some irrationality. Perhaps it is in this respect she considers her view to be practical and concerned with ‘real life’? Rationality and morality is in her view intimately connected. It is in this conclusion I disagree because it relates to egoism in too strict a sense, and seems to consider private what is not really private. Further it seems to assume that language to be necessarily normative must relate to moral beings, which I find implausible.

A second problem is that I cannot see that this is coherent with other aspects of her view. If the categorical imperative is to be different from the Moral Law, this has to allow for some actions to be rational, meaning as being governed by the categorical imperative, without being moral which means that there are proper reasons for performing these actions, reasons that are not moral reasons. Either way, this interpretation is vitally different from the common understanding of Kant. To him, reasons are necessarily shared, egoism is not possible. But why has she chosen to argue so differently from Kant, but yet seems to relate to his strong demands concerning issues like determinism and egoism? One reason I assume is because of the problems

³² Korsgaard 1996a, p. 125.

she appears to find in his arguments based on his conception of rationality. Another, but related concern is that her view is accordingly more practical, aimed at dealing with our actual situations of choice and at the same time she somehow wish to take the ideal aspects into consideration.

1.7 Double- versus single-level theories

Kant is often criticized as having an ideal theory not very suitable for dealing with real life moral situations. One such situation Kant's views are often considered not to be able to meet is how one can deal with others having evil ends without also accepting consequentialism. I suppose Korsgaard wishes to present a position capable of meeting this problem, but without giving up some of Kant's central ideas. As this example of the murder shows, it is not always our duty to tell the truth:

“Deadly knowledge: You ask me whether Grey committed some murder. I know that unless I tell you a lie, you would come to believe truly that Grey is the murderer. Since you could not conceal that belief from Grey, he would then, to protect himself, murder you as well.”(Parfit, 2002, p.288)

Here, it seems hard to defend what would be the Kantian position - that one do what one ought to do by telling the truth, and that the bad consequences are not one's responsibility. Korsgaard says “it is permissible to lie to a deceiver in order to counter the deception.”(Korsgaard 1996b, p. 145) She further refers to what she calls the perfect duty of virtue, that one will not for the sake of humanity be the tool for the devil and his evil ends. However, to argue like this requires, according to Korsgaard, a special structure of our ethical theory. Korsgaard (Korsgaard 1996b, p.147) claims to be able to meet this by unlike Kant allowing for what she calls double-level theories. She follows an idea proposed by Rawls, that we may divide moral philosophy into ideal and non-ideal theories. In an ideal theory everybody is considered to be rational, whereas non-ideal theories have to deal with conditions where the realization of an ideal is impossible. In such situations our aim is not to achieve the impossible, to reach the ideal, but rather be concerned with ‘special principles’ with regard to the circumstances. To be able to deal with both of these, characterizes double-level theories. Single-level

theories cannot distinguish between how we should behave in ideal versus non-ideal situations.

As an example of a single-level theory Korsgaard mentions Kant's view. To him we are always to behave as if we were members of a Kingdom of Ends, this regardless of bad consequences such as in the murder case. The main reason to favour double-level theories according to Korsgaard is that they can provide us with a common sense understanding of responsibility usable for our daily life choices. This allows us to depart from our ideals, principles and standards when these are obviously bad. In other words, Korsgaard argues for a Kantianism that under some circumstances, allows us not to follow the Formula of Humanity and the idea of a Kingdom of Ends, to see them as a "goal to seek rather than an ideal to live up to". (Korsgaard 1996b, p.153) But even this I find to be a too strong demand. It also creates problems concerning how we decide at which level we should operate and when. How can she claim that we are morally obligated, that our moral identity is inescapable, when these kinds of choices are something we need to deal with? In what sense are we morally obligated by necessity? However, what we cannot ignore she claims is the commands from the Formula of Universal Law telling us what we must not under any circumstances do. Hence this law must serve as the ultimate source of justification, Korsgaard concludes. I believe this can be seen as an argument for yet one vital way Korsgaard's view appears to depart from that of Kant: That rational agency is not always the same as moral agency. But for some inexplicable reason Korsgaard apparently does not mean this. So her position appears to be incoherent, or at least not very consistent when explaining central concepts like rational agency and its relation to moral agency. Incoherence's possibly due to how these different levels relate to each other. By arguing for a double-level theory rather than a single-level theory Korsgaard takes one step away from Kant. The question is: Is it far enough?

1.8 Conclusion

Korsgaard's claim that the categorical imperative is not the Moral Law appears to create some challenges. These challenges I find these challenges concern the relation between rational agency and moral agency and believe that this uncertainty is due to a confusion concerning her view on rationality. By claiming that morality is grounded in human

nature, this confusion appears to be to what extent rational nature and human nature is to be understood as the same. Intuitively there seems to be a problem with what Korsgaard wishes to achieve, namely to somehow combine the views of Hume and Kant. For how can one argue for both the necessity Kant associates with moral obligations and the categorical imperative and, at the same time, claim that this necessity is also captured if one sees human agency and reflection as being the creators of morality and of the norms which make claims on us. Who is the author of the laws that bind us, and where does necessity come from if we do not ground it in something transcending human nature? Do they come from the laws of nature? In the next chapters I will explore these concerns further.

2.0 Justification and explanation

In Korsgaard's arguments from the categorical imperative to the Moral Law, there seems to be some common structural features in how she justifies different concepts and principles. In this chapter I wish to shed light on these procedural aspects of her position and, in particular discuss whether or not these procedures manage to capture the necessity she requires from moral obligations and from the categorical imperative. First, I will locate Korsgaard's position in the philosophical landscape as a result of the procedural aspects of her thoughts. Then, I will challenge Korsgaard's view by exploring some ideas related to her provided by Rawls and Scanlon.

2.1 Coherence theories

Justification can be understood as providing good and sufficient reason(s) for having certain beliefs and emotions, for making certain claims, for accepting certain principles, for making certain judgements, for having certain laws, for holding certain views, and for doing certain actions. Thus justified beliefs, actions, emotions, for example, have in common with knowledge of some proposition that both rest upon that we have reasons such as experience or evidence to support them.

In the paper "The emergence of justification in ethics" Dagfinn Føllesdal outlines the history of justification, focusing on and defending the view commonly known as coherentism. I will now use his article as my main source when exploring some aspects of the history of justification. Føllesdal writes that Aristotle distinguished between three different approaches of justification: linear argumentation from a basis, infinite regress and circular argumentation. Aristotle claimed that only moving from first principles in a linear manner would provide justification for what we believe to know. One of the main challenges to this understanding of justification is to find suitable first principles. The two other options, moving backwards into an infinite regress and moving in a circle can not provide such justification according to Aristotle. In the epistemic case, the problem of infinite regress may arise when we want to justify our belief in one claim by the belief of another claim which, again, is in need of

justification - and so on. This kind of epistemic regress problem gives rise to the sceptical claim that justification is impossible³³.

This problem, in the spirit of Aristotle, can also be expressed more as a metaphysical problem occurring if we try to base our knowledge of one matter on the knowledge of something else, without being able to trace the process to an unconditional starting point. However, a different approach, following the thoughts of Plato emerged, based on the idea that we test hypotheses against observations. This may be seen as a predecessor of a more circular approach³⁴, an ‘empirical’ approach based upon agreement or coherence between the general and the particular, as our source of justification.

Defenders of this view include Middle Age philosophers, such as Thomas Aquinas, holding that this approach is best for justification in sciences while Aristotle’s method should be restricted to metaphysics. He further argues that neither mathematics nor ethics seems to be based upon empirical observation, but rather rests on obvious first principles. In 1843, John Stuart Mill opposed this understanding of mathematics, claiming that even mathematics is being tested against observations. However, such observations opposing any principles of mathematics must be met by a higher degree of critical evaluation since changing any principle in mathematics has consequences for the rest of the mathematical system. Furthermore, mathematics is already tested to such a vast degree in our daily life activities, that one should, according to this view, rather question one’s observations than question the principles of mathematics. More recent philosophers, such as W. V. Quine, Morton White and Nelson Goodman develop these ideas further. In the article ‘Two Dogmas of Empiricism’ Quine says that:

“...total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field....Having re-evaluated one statement we must re-evaluate some others, which may be statements logically connected with the first or may be statements of logical connections themselves...No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.”(Quine 1953, pp.42-43)

³³ <http://plato.stanford.edu/entries/justep-coherence/>

³⁴ Føllesdal 2005, “The emergence of justification in ethics”.

Thus our beliefs form a system, where the whole system changes if just one of the beliefs is being modified. Justification of beliefs is hence a system or holistic project. Quine applied this method of equilibrium to science, mathematics and logic. Later on the method of equilibrium was also applied in ethics by, among others, Israel Scheffler and John Rawls. Both Rawls and Scanlon can be said to use versions of this method in their theories of justification. By 'reflective equilibrium' Rawls can be understood as meaning a method of agreement. But this is not his only idea of justification. Also his idea of a public reason and his derivation of principles in the original position may be seen as methods of justification³⁵. (Scanlon p. 139) At first glance it seems plausible that Korsgaard's ideas of reflection and endorsement may be analogous to these ideas of agreement, equilibrium and coherence. In order to place Korsgaard's view on justification in this landscape, I will challenge her ideas with the ones of Rawls and Scanlon.

2.2 Procedural realism

Kant, Rawls, Scanlon and Korsgaard are commonly interpreted as constructivists. In the essay "Realism and Constructivism in twentieth-century moral philosophy" (Korsgaard 2003) Korsgaard argues for constructivism explaining the difference between realism and constructivism as mainly depending on the perspective which we take. Korsgaard concludes her paper by drawing the distinction between realism and constructivism as being not about which one is true or right, but as taking different standpoints towards practical philosophy. The realist view is taking the theoretical standpoint when looking for ethical knowledge in order to apply it to a practical problem. The standpoint of constructivism is practical, aimed at solving practical problems. Somehow, Korsgaard comes to the conclusion that both realism and constructivism are true, I will now try to find out why she thinks this is the case. Where realism is concerned about the nature of a concept, constructivism can be understood as taking a concept as dealing with the answer to a practical problem.

She traces the debate back to Hobbes, who according to Korsgaard, did exactly this - considered morality as a solution to a problem where the problem in question considers how it would be without morality. As a reply to Nagel's conception of

³⁵ This in accordance with what Scanlon writes in Freeman's Cambridge Companion to Rawls, p. 139.

substantial realism, Korsgaard introduces the distinction of substantial versus procedural realism. This distinction is also inspired by Rawls's distinction between proceduralist and substantive conceptions of justice. By procedural realism, Korsgaard means that there are right and wrong ways to answer moral questions as opposed to substantial realism which she claims is concerned about "moral facts or truths."³⁶ Korsgaard's main argument for procedural realism is that the opposite would mean that there are no reasons or ought's. She continues that the problem with substantial realism is because it implies that things have normative properties, that there exist moral entities. So in a sense, what there is, is what we construct - a procedure. In this sense realism and constructivism is the same. This basic idea seems to be central to Korsgaard's position. The emphasis she places on reflection, as the creator of conceptions such as morality and identity, suggests that procedural realism and constructivism are defining for her view.

Others, like Scanlon, may also to some degree support these ideas, but not necessarily to the same extent. However, Korsgaard and Scanlon disagree in how deep constructivism can go. To Scanlon the notion of being a reason is primitive, which means reasons cannot be constructed. He can be viewed as a constructivist with regards to moral principles, principles we construct by asking which principles cannot be reasonably rejected. That what has normative force on us is not a belief about an external world, but rather the procedure forcing our conclusion about what are our reasons to have certain attitudes. In this sense he argues in accordance with Korsgaard's idea of procedural realism. Korsgaard takes constructivism further and appears to agree with what is commonly considered the Kantian idea that all reasons can be constructed. Someone like Rawls, on the other hand, according to Korsgaard³⁷ constructs the principle of justice from the original position, while taking the notions of 'best' and 'good' as irreducibly normative. Thus an interesting question is then how far should constructivism go? I will not try to answer this question, but instead focus on Korsgaard's constructivism exploring my intuition that she goes too far.

2.3 Original position argumentation

³⁶ Korsgaard 1996a, p.35.

³⁷ Korsgaard in "Realism and Constructivism in Twentieth-Century Moral Philosophy".

John Rawls's 'original position'³⁸, where what is at stake is to make a rational choice based upon a certain kind of lack of information or restrictions put on our reasoning, or as he puts it – from a veil of ignorance. The idea of a veil of ignorance could be interpreted as a standpoint of choice where one is not motivated by any personal concerns or knowledge. The original position is an initial one securing a fair fundamental agreement. In, *a Theory of Justice*, Rawls aims to outline a theory of justice establishing the idea of fair cooperation involving free, moral and equally worthy agents. It is a society where everyone acknowledges the same ideas of what is meant by justice, and agrees upon which social institutions best satisfy these principles of justice. Hence he argues for a common understanding, or a public reason capable of justifying claims about institutions in that society. Rawls does however, not believe in a common social conception of what is to be understood as the good. Therefore in developing his theory of justice, he cannot derive the concept of right from such a common understanding of the good. Hence instead of arguing in a teleological manner, Rawls argues in a procedural or original position manner, what is to be considered right is what would hypothetically be considered so from the standpoint of the original position. This means a sort of reasoning starting from the original position - reasoning based upon a certain lack of knowledge. Rawls base his conception of justice upon what he calls *The Kantian Interpretation of Justice and Fairness*, an interpretation of Kant's notion of autonomous action as meaning an action performed by someone with a "nature as a free and equal rational being"(Rawls 1999, p.222) Further Rawls argues that the principle of justice can be compared to the categorical imperative since both are principles for someone due to their nature as free and equal rational beings, and the original position can be understood as an interpretation of this virtue.

2.4 Justification of the categorical imperative

Korsgaard compares the structure of two problems as stated by Rawls and Kant. Rawls's problem is a paradox easily explained if imagining liberal politics in a nation

³⁸ What I write about original position and reflective equilibrium is an interpretation primarily based upon reading Rawls *a Theory of Justice* and Scanlon, Rawls on Justification, in Samuel Freeman, 2002a, "The Cambridge Companion to Rawls", Samuel Freeman, "Rawls", and Korsgaard "Reclaiming the History of Ethics: Essays for John Rawls".

which is not liberal; where the problem is caused by the central liberal idea that the people governed by the politics should also agree to the principles of that politics. Rawls's (and Hobbes's) solution to this problem is that of justice.

A similar problem arises when Kant explains the foundation of the categorical imperative³⁹. The problem arises due to the definition of a free will as a sort of causality which is not itself affected by any external causes, and Kant's explanation of this phenomenon as due to the autonomy of the will – meaning that the will has its own laws. But how does the will decide which principles or laws to follow? If the answer to this lies outside the will, the will is not free, but if the will does not yet have any laws – how can it make any decisions? It seems that there must be some restrictions put on the free will. Korsgaard continues to argue that this problem and Rawls problem have the same structure in the sense that:

“In both cases what we are looking for is principles themselves, for we need reasons, ways of choosing and justifying our actions or our policies, and reasons are derived from principles. Yet the very structure of the situation seems to forbid us to choose any particular principles.”(Korsgaard 2003, p. 114)

She continues that Rawls's and Kant's solutions take similar forms; for Kant the solution is the categorical imperative – the law of the free will. The free will is its own law, and that is all it is – according to Korsgaard. Thus the categorical imperative does not have content, it only tells us to choose: “it must choose as its maxim that it can regard as a law”. Rawls's solution is similarly given in by that his two principles of justice describes:

“what a liberal society must do in order to be a liberal society, just as Kant's principle describes what a free will must do in order to be a free will. Rawls's principles are derived from the idea of liberalism itself, in the same way Kant's categorical imperative is derived from the idea of free volition.”(Korsgaard 2003, p. 115)

Both Kant and Rawls are constructivists, she argues, in the sense that their practical concerns are not about getting knowledge to apply in practical situations, but to use practical reason to solve practical problems. In a *Theory of Justice* Korsgaard interprets Rawls as distinguishing between the notions of a concept and of conception as where

³⁹ Korsgaard 2003, p. 114.

'concept' is referring to the solution to the distribution problem and where the conception of justice is the solution to this distribution problem. The same structure, she continues, is to be found in Kant's theory. Here the problem is that of freedom: by reflecting on that problem one arrives at the solution which is the categorical imperative. By reflecting about how a free will must deliberate we arrive at the categorical imperative. The free will deliberates by choosing a law for itself. So the categorical imperative is constitutive of deliberation and at the same time something we arrives at by deliberation. To me this sounds problematic. She continues that the categorical imperative is the logic of practical deliberation, much in the same way Rawls principles can be viewed as the logic of liberalism. (Korsgaard 2003, p. 115) But this I find to be far less problematic for Rawls than it is for Korsgaard since it does not give rise to the same kind of circularity as it does for Korsgaard's argument. Thus, she concludes, according to constructivism normative concepts are not names of facts or objects in an external world, but names of the solutions to a problem.

Is this in her view restricted only to relating to moral concepts, since only they are about practical problems? She does not answer this question when the concepts at stake are those of science, but argues that some more trivial concepts from our daily life may also have this structure. Another important aspect of constructivist theories is how far they go. As an example Korsgaard mentions Scanlon, who when constructing moral principles, wants to solve the problem of justifiability. The question he wants to answer is which principles one can reasonably reject. However, as we recall, he stops his constructivism here, claiming that the notion of being a reason is primitive, not something which we construct. Presenting a special form of contractualism, he argues that it is the notions of right and wrong, or of "what we owe to each other", which is most important. The idea of justifiability serves both as a basis for the morality of right and wrong, and as a characterization of its content. He addresses the issue of deciding between right and wrong as being a matter of finding which reasons or principles no one could reasonably reject.

To Korsgaard constructivism goes a bit further. Reasons are not primitive, but are derived from principles she claims. However I find it problematic that deliberation, if rational, is to deliberate in accordance with the categorical imperative. How then can the categorical imperative be justified by deliberation if it as well is constitutive of

deliberation? Practical reasoning should in Korsgaard's view presuppose the categorical imperative. Is it somehow established in a non-rational manner, or is constructivism only concerned about our awareness of it, that it is a construction for us, but that it is a principle which somehow existed before we had this awareness? But does this not imply that what we have is a view not based upon procedural realism as Korsgaard claims?

2.5 Reflective endorsement - the procedure of the categorical imperative

How do we arrive at the original position? The very structure of an original position argument is justified by the method of reflective equilibrium. Scanlon suggests that when applying Rawls's method one may proceed as follows: First one identifies, in an impartial and fully informed manner, a set of judgements about justice to be adequate and correct. Then one tries to formulate principles coherent with these judgements. Finally one evaluates whether or not this coherence is successful, and if not, one will have to decide which to give up – the principles or the judgements. This moving back and forth between principles and judgments continues until there are no more apparent conflicts between them. When this is the case, which it never is since it is an ideal state, we have reached a 'reflective equilibrium'. However, this process is Socratic in the sense that even though we have a perfect match between our judgements and their principles, our knowledge of these principles may want us to continue our search. This process is one we do ourselves, from a first person perspective and is, according to Scanlon, "a process of deciding what to think, not merely one of describing what we do think." (Freeman 2002a, p. 149). The method of reflective equilibrium is a process aiming at making good decisions, even though a perfect choice is not a realistic end point. Thus, Scanlon concludes that the main aim of the method is one of deliberation, but this is not in conflict with a third person descriptive standpoint. Rather, our ability to see the process in both a deliberate and descriptive manner is vital owing to the fact that the procedure of 'reflective equilibrium' having different stages – benefits from different perspectives. Scanlon suggests calling this process "reflective modifications of ones reasons". This reasoning or deliberation arriving at the original position is not motivated by an idea of justice; a conception of justice rather emerges by the procedure of this deliberation. The process of arriving at the original position may be viewed as a

matter of reaching what Rawls refers to as a reflective equilibrium, a consensus in our community concerning our judgements about what are reasonable principles of justice. This equilibrium between people's conceptions of what is meant with justice hence decides the veil of ignorance as to what is irrelevant for justice. This ignorance ensures impartiality despite people making choices based on self-interest. Rawls applies the method of reflective equilibrium both in the justification and in the construction of his theory⁴⁰.

Perhaps it is something similar to this Korsgaard too has in mind, that the categorical imperative is something we become aware of through deliberation? However, I do find this more problematic than Rawls's emergence of the idea of justice, since it is not as fundamental to the very procedure from which it is being constructed. Not only in establishing central concepts of her theory like that of the categorical imperative does Korsgaard base her view on deliberation, but also for the more practical concern of deciding our reasons for action. Where the agreement in question is what she calls endorsement, I assume it can be viewed as some sort of internal agreement. This internal agreement I have argued does not secure that these reasons are also reasons for others. Where Rawls talk of our nature as free and equal rational beings, Korsgaard does as we recall speak of human nature. Her distinction between the Moral Law and the categorical imperative (a distinction I cannot see that Rawls makes) requires some extra arguments, arguments I have argued don't necessarily do their job. But I believe Korsgaard would disagree because as she says; reflective endorsement is a test for normativity, a test all of our inclinations must pass in order for them to be considered reasons. Actually, she continues, it is a test for all of our motives and inclinations. The normative question is the one of whether our nature is of the good and if we should act according to its claims. It is a question asking for the very authorization morality has on us. However, she continues, normativity is a problem for us because of our reflective nature. We can always question our beliefs and motives: we can question the claims normativity makes on us. However, if our moral beliefs and motives sustain the test of reflection, what we have found out according to Korsgaard, is not about something else such as the existence of alien normative entities, but that this process in itself is normativity. In this sense morality is constructed by our reflection. The method of

⁴⁰According to Scanlon, in *Rawls on Justification*, Samuel Freeman, 2002a.

reflective endorsement is hence a method which provides a test, the test of reflection due to the categorical imperative. Korsgaard takes these ideas as her starting point when arguing that the solution to a problem lies in a good formulation of the problem. Thus by questioning morality and normativity, we will, by reflection, find that the very procedure is the answer to our question. The reflective structure of our mind makes it necessary for us to decide between our impulses, we are forced to decide which reasons we have to act upon. It is a test not only justifying morality, *it is morality itself*⁴¹. Thus in her view, morality is constructed by reflection. Further and related, her conception of value has a similar structure.

“Value is grounded in rational nature – in particular in the structure of reflective consciousness – and it is projected on to the world.”(Korsgaard 1996b, p.116)

Hence, also value is something we create and, in her view, is not something which we find out there. Values are due to the procedure of making laws to yourselves. We create value by willing a maxim to be a law, and it is this Korsgaard understands with procedural realism – our constructions of values. To will a maxim to be a law requires someone who wills it - it must be a reason for someone. This identity is something we have by being self-conscious. It is by valuing yourself, that you get a conception of yourself. This conception is not a theoretical one, but rather an expression of a practical identity: an identity based upon that you value yourself, and find your reasons and your life worth having and living. According to Korsgaard, examples of such identities are that you are a human, that you belong to a nation and that you are a woman and so on. The central point is that these practical identities are something we attain by reflection; they are also in a sense procedural. Even moral obligations emerge through reflection. That both values, obligations, reasons, our identity and so forth are all constructed by similar procedures makes one worry about the danger of becoming self-referential or circular. But perhaps the practical identity in question is not the only conception of identity that Korsgaard presents? Or does she also operate with a more theoretical, or pre-reflective abstract notion of unity or of the self more in accordance with Kant and his ‘transcendental unity of apperception’⁴²? A notion of the self which could prevent

⁴¹ Korsgaard 1996a, p. 89.

⁴² I will refer to this term as used by Allison in “Idealism and Freedom”, “Kant’s theory of freedom” and “Kant’s transcendental idealism”.

what appears to be a circular view basing all central concepts on a similar procedure? I will explore her conception of identity a bit further in order to see if she can meet these concerns.

2.6 Personal identity and unity of agency

In the paper “Personal Identity and the Unity of Agency: A Kantian Response to Parfit”, Korsgaard argues against Derek Parfit’s conception of a person or agency as first of all a form of experience. Korsgaard instead advances the idea that we primarily view ourselves as agents, as something which we do. Parfit gives an account of what he refers to as the standard view of personal identity. This is the view that “the persistence of an object over time can be understood in terms of spatiotemporal continuity under a concept” (Korsgaard 1996b, p 364) where a person is, or coincides with, exactly such an spatiotemporal continuum. But, Korsgaard continues, this view cannot deal with the fact that human beings may change radically and hence break this physical continuum, but still have what Parfit calls psychological continuity: a continuity expressed by, for example our tastes, emotions, characters and memories. With psychological continuity Parfit means, according to Korsgaard, that there must be some kind of causal connection between our psychological states. To obtain this kind of connectedness Parfit considers two possibilities: the reductionist view where the conception of a person is reduced to something physical “the existence of a brain and body, and the occurrences of a series of interrelated physical and mental events”⁴³, or on the other hand the non-reductionist view where the continuous subject is a “deep further fact” about someone. He claims according to Korsgaard that the important aspect of a reductionist view is not so much about the material aspects of a person but rather in how matter is organized. So Korsgaard continues, with identity he means something like:

“what we normally count as persisting identity is simply formal continuity plus uniqueness-that is, being the only formal continuer of a past self.”(Korsgaard 1996b, p. 367)

But, what is important is according to this interpretation of Parfit is not our personal identity, but what Parfit denotes ““Relation-R” – psychological connectedness and

⁴³ Korsgaard 1996b, p. 368, quoting Derek Parfit 1984, in *Reasons and Persons* p. 211.

continuity” (Korsgaard 1996b, p. 366) The reason why he according to Korsgaard think “Relation-R” is more important than personal identity is due to “Relation-R” being copiable. What is essential to a person is what can be copied, it’s that persons formal aspects. It is not even sure that this relation establishes any connections between a person and whoever will occupy his body in the future. Korsgaard argues against this idea that we do not necessarily are identical with whom occupies our body in the future. She explains that we do not need to give a metaphysical argument for this, since our reasons for seeing ourselves as one and the same rational agent over a period of time is purely practical. The unified self-conception which Korsgaard has in mind is a practical unity, a unity depending on our standpoint of deliberation and choice.

What is important is to ensure that there are no conflicts among your motives. To ensure this and, as well, support the reductionist view, Parfit refers to the example of split brains. In short, this is a problem occurring when the nerves between the two hemispheres of the brain are cut so that the two hemispheres operate independently. This means that one part of the brain can operate without being conscious of the activities of another part of the brain. Thus from a metaphysical point of view the two hemispheres are no closer related than any two human beings could be. However, since there is only one body to perform an action, the two parts will need to come to some sort of agreement.

“The unity of consciousness consists in one’s ability to coordinate and integrate conscious activities. People with split brains cannot integrate these activities in the same way they could before.”(Korsgaard 1996b, p. 376)

In split brain cases Korsgaard argues that all that is needed is some communication between the two hemispheres, and in order to communicate one does not need a common psychological subject. What is needed is the unity of agency she continues. I assume this unity of agency must then be that one of agreement: the agreement reached by deliberation and discussion between the two hemispheres, ending in a conclusion on how to act. She compares this with the unanimous decision the parties in Rawls’ original position must come to. But who is to judge this unity as such? Is it from a third person point of view or from a first person point of view? If it is from a first person point of view, how can one ensure that it is not only one of the hemispheres who do the

job and make the decisions and have the feeling of unity? Her conception of personal and practical identity rests heavily upon what she calls a Kantian view: the view that we may see ourselves from two different standpoints. This interpretation Korsgaard claims to find in Kant's writings several places and in particular in *Foundations of the Metaphysics of Morals*. However, this view on unity and personal identity also seems to diverge from Kant and his idea of the 'transcendental unity of apperception'. What Kant means by this is widely discussed, a discussion I will not become involved in here. However, I do think it is uncontroversial to suggest that by 'transcendental unity' Kant had in mind a unity as a necessary condition for us to be able to be conscious and to have any cognition at all. Korsgaard's idea of unity seems to be the opposite, that of a unity resting on agency and emerging through conscious activities. Although it is difficult to see how the transcendental unity can explain split brain cases, so in this respect, Korsgaard's suggestion may be superior. One could also suggest that Korsgaard's idea of a unity of agency can be combined with the idea of a transcendental unity, and that the latter is a necessary condition for each hemisphere to deliberate in the first place. Hence from this I find one cannot conclude that Korsgaard argues for a view dismissing the transcendental aspects of Kant's philosophy, but only that it is not to what she pays attention⁴⁴. To explore these issues further I will now see what she understands with two standpoints.

2.7 Two standpoints and the Modern Scientific World View

To solve the apparent conflict between freedom and natural laws, Kant introduces the distinction between the noumenal (things in themselves) and the phenomenal (things as they appear) worlds. Kant's noumena/phenomena distinction has given rise to several interpretations. 'Transcendental idealism' is commonly understood as only seeing things as they appear to us and not as they are in themselves. This combined with the idea of empirical realism is often denoted as a 'two-world' theory. Another interpretation is the 'two-aspect'⁴⁵ account given by Henry E. Allison. His 'two-aspect' account is based upon his conception *transcendental idealism*, as argued in his book

⁴⁴ I will discuss in chapter 4 whether or not some of her arguments from the categorical imperative to the Moral Law are transcendental. Her conception of unity of agency is in this respect important.

⁴⁵ In "Two Standpoints and the Belief in Freedom" Dana K. Nelkin describes what she claims to be another version of the two aspect account, namely Donald Davidson's anomalous monism as described in "Mental Events".

Kant's Transcendental Idealism. Allison here argues for a view opposing the standard interpretation of Kant's transcendental idealism as a metaphysical theory concerned with what is knowable or not. Instead of this he argues that transcendental idealism is a view based upon seeing the distinction of noumena/phenomena as distinguishing between different aspects of a world rather than dealing with two worlds.

As a third alternative, Korsgaard argues for the 'two-standpoint account', an account one of course can question how relates to some versions of transcendental idealism⁴⁶. However this is not what I am about to discuss. Now I will explore what she means by a 'two-standpoint account', and start with seeing which questions inspired this distinction.

According to Korsgaard, Hobbes and Pufendorf asked how the mechanical world of nature could include moral properties. This way of questioning is due to the fact that they had adopted what she calls 'The Modern Scientific World View', that they believe it takes something like God or a sovereign to break the order of nature. The Modern Scientific World View is the view which according to Korsgaard, is modern in the sense that the world is not following the antique idea of world as form, and that what is real is what is good. Rather Korsgaard argues⁴⁷, form is forced onto matter, and it is obligation which forces value. According to Korsgaard, Aristotle argues that when we seek excellence we do so because it seems attractive to us. Korsgaard opposes this and rather follows the Kantian idea of obligation as a compulsory force. She quotes Kant and says that "reason – which is form – isn't in the world, but is something that we impose upon it." (Korsgaard 1996a p. 5) She continues to argue that the idea of autonomy as the ethics of obligation is the only one capable of dealing with the modern world. The Modern Scientific World View she opposes to the Ancient Greek World View and Medieval Christian World View, and so one begins to suspect that the view she proposes is not so modern. This suspicion deepens when she advances our challenge to our freedom as being determinism.

I believe by determinism she means in the causal sense as seeing every event as caused by a prior event, and that what will happen in the future is fixed because of some laws. It was natural for Kant to hold this view living at the time when Newton was the authority. However, in science today, thanks to research in several fields, it is highly

⁴⁶ Related questions will be discussed later in the thesis.

⁴⁷ Korsgaard 1996a, p. 4.

questionable whether determinism is true. In quantum mechanics, relativistic theories and theories concerning singularities such as black holes, the view of determinism is challenged. In many scientific fields today the idea of probabilities better explains the phenomena investigated than the idea of determinism. However, I will leave this discussion, and rather explore further Korsgaard's view on how to deal with her ideas of freedom and how it can be combined with the Modern Scientific World View. Her solution echoes Kant by emphasizing the distinction of theoretical and practical reasoning, as a distinction between perceiving and doing. She continues to argue that in order to make rational choices, we must value what we pursue. Further, we must value ourselves, since we choose to do what matters to us: hence a theory of value – by Kant stated in his Formula of Humanity – that human beings must be considered as “ends-in-themselves”. Human beings have value in themselves. Therefore, when making a rational choice, we must consider ourselves as citizens in “the Kingdom of Ends”, and only act in a way which shows respect for ourselves and other people. We must only act in a way that may serve as a universal law. Since it is because we are active we are valuable, Korsgaard here argues against the common understanding of Kant's distinction between the noumenal and the phenomenal world as a metaphysical or ontological distinction, as if we were living in two different worlds simultaneously. Rather, Korsgaard suggests an interpretation that there is one world which we perceive in two different ways, from two different standpoints. The important distinction according to this view is the passive versus active, and not one concerning which world is most real.

By viewing practical reason as an active faculty of mind, one also takes the important questions concerning one's identity not to be metaphysical ones, but rather practical. So it is a conception of identity being practical in the sense of being a cause – rather than something being caused. And it is due to this that we are free and can be held responsible for our actions. Korsgaard, however, claims that some problems we have concerning responsibility are due to our failure to acclaim Kant's radical distinction between theoretical and practical reason and their domains of explanation versus deliberation. Kant's distinction between practical and theoretical reason is a distinction commonly understood as an ontological or metaphysical one, as a distinction concerning what is ‘most real’. Korsgaard suggests another interpretation, arguing that

the division of our mental faculties allows us to attain two different perspectives or standpoints from which we can relate to our actions. Theoretical reason's territory is, in this view, the phenomenal world – the objects as we experience them – as we passively receive them. This is what we may call the world of knowledge or the world of scientific explanation. The domain of practical reason is the 'noumenal' world – things as they are in themselves. Thus, she argues, we have two standpoints allowing us to be both deliberate choosers on how to act and, at the same time being capable of giving causal explanations of our behaviour. What she calls *the Scientific World View* is the view serving the purpose of predicting and explaining. It is as such the domain of knowledge and truth. Korsgaard describes the 'noumenal' world as the ethical world, the world in which we are active beings and the creator of our thoughts and actions. It is in this sense that we have value, as active beings and choosers – as noumena. Hence Korsgaard clearly distinguishes the issue of deciding what we ought to do, and to give an explanation of why we ought to do something, as a distinction being a matter of perspective. When asking for justification of our actions we take the first person perspective, focusing on what it is that makes a claim on us we look for support for our normative criterions. On the other hand, when wishing to explain why we should do something, we do that from a third-person perspective. The first person perspective is rather something we construct and not a matter of what we perceive.

In what way does her view depart from Allison's 'two aspect' view? To Allison a rational character has both an intelligible and empirical character. Thus it seems that the aspects in question are aspects of rational beings as seen from a third person perspective, and not two perspectives as experienced by the person in question which is Korsgaard's view. From a first person perspective it must be agreement between the two parts of the brain that unifies the agent, not the action. So to me it seems that Korsgaard's argument in the split brain case is problematic, and that what she, in fact, argues for is a version of a 'two aspect' view and not the 'two perspective' view since she is dependent on a third person perspective.

2.8 Pre-reflective intuitions⁴⁸ and the unconditional

⁴⁸ Inspired by Dagfinn Føllesdal's paper "The emergence of justification in ethics".

A common critique of the method of reflective equilibrium according to Scanlon, is the one of relativism, that where our process ends is decided by the set of judgements from where we started. Scanlon argues against this by saying that the set of judgements from where we start our reflections is not a fixed set; this argument does not remove the problem completely but it makes it smaller. Another problem is that such a process, by the nature of the process, has certain pre-implied ideas in it. How is it possible to talk of justification without already accepting the idea of responsibility? If we cannot make choices, if we are not responsible - how can we justify anything – then justification is not in our powers. The idea of responsibility seems to be presupposed in the very idea of justification. Morality may be seen as putting constraints to our rational deliberation towards ends.

Rawls's idea of 'reflective equilibrium' is influenced by Nelson Goodman's 'empirical' method for testing particular conclusions with general principles commonly accepted, and vice versa, testing general principles to particular conclusions commonly accepted. Through a process of balanced modifications one arrives at an equilibrium providing a justification (at least temporarily) of induction.

“How do we justify a *deduction*? Plainly by showing that it conforms to the general rules of deductive inference..... But how is the validity of rules to be determined? Here again we encounter philosophers who insist that these rules follow from some self-evident axiom, and others who try to show that the rules are grounded in the very nature of the human mind. I think the answer lies much nearer the surface. Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences...This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that the general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.*”(Goodman 1983, pp. 63-64)

Justification thus does not lie in the rules or in accepted inferences, but rather in the equilibrium between them as the source of agreement. Goodman sheds light on an important feature of the reflective equilibrium method⁴⁹, the existence of pre-reflective

⁴⁹ Føllesdal 2005, p. 177.

intuitive acceptances of some statements. The point is that when we use this method we do not, at least not at the same time, question all of our judgments. So what is this intuitive acceptance? Rawls suggests that one such main source of evidence is perception, not only in sciences but also in ethics. Some particular moral judgements are privileged in that they serve as evidence for other general ethical principles. This is in much the same way as in science where particular observations modify general hypotheses.

The problem of circularity according to this view is not a problem given that one accepts some sort of presuppositions, challenging the circle. To Korsgaard what could these presuppositions be that should help her to avoid circularity? As we recall she can be understood as operating with two kinds of unconditional; the good and rational will and the idea of us as reflective conscious beings. It seems that her pre-reflective intuition that we are responsible for our action combined with seeing determinism as a threat to this quest, have lead Korsgaard to aim for an unnecessary strong foundation of the norms binding us. But one problem with these pre-reflective ideas is that they cannot be questioned. This makes a vital difference from the reflective equilibrium method where all judgments in principle, at some point can be questioned.

Kant addressed the problem of the free will due to the threats of determinism, and he tried to solve the apparent dilemma between determinism and freedom by appealing to several dualities. Rawls' means to have a theory that don't need to make claims of the deep dualisms he means to find in Kant. The method of reflective equilibrium is the basis for an empirical theory⁵⁰, a theory based upon a reflective process between our particular observations and our general principles of justice, concluding in equilibrium between the two.

Korsgaard argues for a method having the same structure, but where what one seeks is not a theory of justice, but rather to justify the claims normativity makes on us. But where Rawls seems to argue for a procedure in which both our general principles and our observations are being challenged, for Korsgaard the procedure appears to go only one way – reflective endorsement is a test to see whether our reasons are in accordance with the categorical imperative. The principles of reason are not to be challenged by this procedure, so the methods are not 'empirical' in the sense Rawls had

⁵⁰ Rawls 1999, pp. 226-227.

in mind where principles may be adjusted by observations. Hence, Korsgaard offers an account arguing for the importance of keeping Kant's dualities, but at the same time argues in accordance with Rawls's proceduralism when establishing the categorical imperative. One problem is whether this is really possible. Similar to Korsgaard and Rawls, Scanlon too argues for a method similar to the one of 'reflective equilibrium' when we justify the reasons we have for doing something. However, unlike the other two Scanlon has given up a central duality – the one between theoretical and practical reasoning. This means that the domain of responsibility is not as it is for Korsgaard only a matter of deciding what we ought to do, it is also a matter of what we ought to believe, think and feel – we are responsible for all our attitudes. Korsgaard appears to aim at similar ideas:

“Concepts like knowledge, beauty, and meaning, as well as virtue and justice, all have a normative dimension, for they tell us what to think, what to like, what to do, and what to be. And it is this force of these normative claims – the right of these concepts to give laws to us – that we want to understand.”(Korsgaard 1996a, p. 9)

Korsgaard has been criticized for focusing too much on moral normativity, and neglecting other forms⁵¹. As I read Korsgaard, the reason why she focuses on the moral form of normativity is because she finds it to be superior to the other kinds, that other kinds of normativity are derived from the normativity in the moral case. But, if it is possible to read her view as I do, that we can have rational normativity without having moral normativity, would not rational normativity be superior? I also believe she encounters problems that Scanlon avoids. If normativity is constructed by a procedure of reflection, how can one then get to the normativity of thinking by thinking? To accomplish this normativity must have its source somewhere other than from our reflections. If she thinks that these norms oblige us, or are forced on us by necessity as in the moral case, what is this necessity based on if not on reason in a Kantian sense – on human nature and the laws of nature? As an answer to these kinds of concerns Korsgaard argues that there is a normative perspective from which morality can be questioned.

⁵¹ Bernard Williams, in “History, Morality, and the Test of Reflection”, Korsgaard 1996a, pp. 210-213.

“It is human nature to be governed by morality, and from every point of view, including its own, morality earns its right to govern us. We have therefore no reason to reject our nature, and can allow it to be a law to us. Human nature, moral government included, is therefore normative, and has authority for us.”(Korsgaard 1996a, p. 66)

But is it not the same to say that, yes we can question morality, we can reflect upon it, but we are by our very nature forced to come to a conclusion in favour of morality. Is this is the essence of reflective endorsement? Then perhaps Korsgaard’s view is better understood in light of what she refers to as Hume’s reflexivity test, than a test in coherence with Kant’s view? This is a test Korsgaard explains to be when the understanding when reflecting on itself and its own procedures ends up in doubts. Contrary to this is morality, where reflection on our own moral sentiments makes those sentiments stronger. Hence scepticism of the understanding is fine, while scepticism of morality is out of place, according to Korsgaard’s reading of Hume, a reading she appears to adopt.

2.9 Conclusion

Where Rawls seems to apply the method of reflective equilibrium both as a foundation and as a justification of his theory, Korsgaard appears to take this one step further. To her, this method is in addition a method dealing with how we actually operate when we are about to make a decision on how to act. Her aim is a very practical one: however, as I have argued, this is not unproblematic. How can we establish the normativity of reflection by a procedure of reflection, or to establish the very principles of reflection by reflection? I suppose Korsgaard would reply to these worries about circularity, by referring to the concepts of autonomy and spontaneity: concepts which, if interpreted in the air of Kant, could be said to break such circularity. But the question then is, is Korsgaard’s conception of these concepts similar enough to Kant’s understanding of them to achieve the same thing? Where Kant wanted to establish objectivity, Korsgaard takes a different approach wanting to consider in a more objective fashion our normative considerations. Again, we are dealing with a procedure, a procedure which, again, depends on another procedure. What is valuable, what is our reason to act, what is moral, which identity we have – all are a result of a procedure – and since all these issues are dependent issues it is hard to see how to avoid problems due to self-reference. Kant avoided this by introducing the idea of the transcendental, and it appears difficult

to see how to avoid problems of self-reference without in some sense introducing some meta perspectives – or at least to base the justification of some of these concepts on something ‘outside’ or transcending the individual person. In the last chapter I will take a more detailed look at this challenge. But first I will investigate whether or not her view of rationality is somewhat similar to Kant’s conception, and to what degree her view contains the dualities that Kant is commonly assumed to support. In particular I will do this by exploring how she relates emotions to rationality in deliberation, to see whether emotions can serve as the needed force for us to be morally obligated.

3.0 Deliberation and the dualities of reason

Since so many of the central ideas in Korsgaard's view are constructed by the procedure of deliberation, in this chapter I will examine some aspects of it further. As we recall, to Korsgaard the deliberate process of justification is something we do from a first person perspective. Explanations, on the other hand, are according to her, from a third person point of view. These perspectives are consequences of a distinct division of our mental faculties, the distinction between practical and theoretical reason. Thus reason is understood as a standpoint for reflection, where practical reason is concerned about what to do and theoretical reason is about explanations, predictions or in what to believe. Central in understanding what Korsgaard means by deliberation, is therefore to understand what she means by third person and first person perspectives. Internal to the procedure of deliberation, there appears to be another division, one between our impulses and of our rational activities. One question is if this distinction is essentially similar to the division of her perspectives. In order to explore this division I will look in more detail at what appears to be a duality in her position, the one between rationality and emotions:

“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 oversimplified, division between natural impulses that do not belong to my proper self and rational impulses that do.”(Korsgaard 1996a, p.240)

So perhaps she has succeeded in this, to make the dualities of Kant less apparent?

3.1 Korsgaard's conception of rationality

Korsgaard argues in accordance with Kant that among our rational capacities is our ability to set an end, to decide what a purpose for us to act is. For such a choice among our reasons to be rational, our choice must be based upon a lack of interest for that end. This is the same as requiring that our choice is not based upon our private reasons, inclinations or desires, but in accordance with a categorical imperative. Thus the categorical imperative appears to be the principle which we must follow in order to act rationally. To Kant this is directly linked to acting morally, since the categorical imperative is the Moral Law. But to Korsgaard the situation is different. To get a clear

understanding of Korsgaard's conception of rationality is not easy, as I find it often to be implicit in her writings, so to give an account would be speculative. Therefore I will restrict myself to a short outline.

Sometimes she gives the impression of sharing the conception of rationality with that associated with Kant, at other times she argues in ways which appears to be incoherent with the very same ideas. Central to what she argues to agree with Kant, in is that the principles of practical reasoning spring from the rational nature of the will, and that such reasoning conclude with what we ought to do.

In the previous chapter I established that there is a problem concerning the justification of the categorical imperative, and about how Korsgaard argues that it should have such a strong demand on us. In this chapter I will see whether related concerns can be established with regards to our deliberations. Korsgaard claims that to be able to account for the reasons which move us is characteristic of rational agency. In this view, our reflective capacities, our capability to be aware of our own reasons, is what distinguishes us from other living creatures. Thus when we act, how do we decide what to do? How do we know which reasons we have for acting in certain ways, and how do we choose between these reasons? To answer these questions Korsgaard presents a view relying heavily on Kant and his idea of an unconditional starting point, the idea that the only thing which is unconditionally valuable is a good – or rational - will. Thus to be rational consists of responding correctly to reasons, but being a reason is not primitive. According to Korsgaard our reasons needs to be in accordance with our rational principles. From a practical point of view, according to Korsgaard, to have reasons means to have reflective success. When we decide that something is a reason to act on, that is a decision we have reached by reflection. It is a reflection consisting of distancing ourselves from our inclinations and questioning them. Thus what dominate deliberation are our rational principles and not our desires or emotions and in this sense her understanding of deliberation has more in common with her understanding of Hume's conception of justification than of Kant's view on reflection. As we recall, Korsgaard explained that to Hume justification emerged from of us adopting both the theoretical and the practical perspectives. Perhaps this inspired Korsgaard's idea of the justification of deliberation as somehow internally to take different perspectives?

With reflection she seems to refer to a kind of reasoning, one based upon certain rules of logic and with the categorical imperative as its principle, I assume. This reasoning is supposed to be completely free from one's inclinations. How is this possible? Even the simplest mathematical reasoning, with a limited set of axioms and rules gives rise to situations of choice - situations where there are several, and sometimes equally good ways to come to an end. So how are choices of this kind to be made? At random, spontaneously or perhaps determined? Or does this kind of reasoning require some additional motivational force, such as emotions? The reflections Korsgaard has in mind I presume to be of an even more complex kind than simple mathematics, and as such constantly give rise to situations of choice. But if one assumes that emotions are necessary for this kind of reflection or procedure to proceed, this too gives rise to some problems. Since then what we have is in a sense a new situation of deliberation internal to the original procedure, which again would require an internal procedure and so on. It seems that this would lead to an infinite regress so there must be some other alternatives.

One such alternative she points at is to consider one self as another 'I'. What does this mean? To consider oneself as another 'I' owing to a small difference of time? This I presume is not an option open to her since, as we recall, our conception of identity is emerging in time and is not something which we have at a particular moment. Perhaps then, this other 'I' is meant in the sense of being as if I was another I. However then one can ask, what is the difference between this other 'I' and a third person? Perhaps it is just a way of taking the third person perspective, to view one's own inclinations objectively. In that case the deliberate perspective becomes very close to the perspective of explanation.

3.2 Scepticism about practical reason

In "Scepticism About Practical Reason"⁵² Korsgaard argues against doubts concerning whether reason can serve as a guide on how to act. She divides the question into two: one is concerning motivational scepticism – whether reason can motivate us; the other form of scepticism she refers to as content scepticism - doubts concerning our principles

⁵² Korsgaard 1996a.

of thought and to what extent they have content capable of helping us in situations of choice and action. Scepticism about practical reason is stated by Hume as following:

“Reason is, an ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”⁵³

What Hume may appear to say is that all reasoning with a motivational force must have started with a passion? Korsgaard argues that if reason plays this limited part, reason cannot decide if our passions and desires are rational or not, then actually there is no way these passions can be rational or irrational. The challenge Korsgaard is then facing is that if one argues as she does, that practical reason may reach conclusions not in any way depending on passions, how can such an operation end up motivating us? Or put differently, can an emotion arise from anything other than an emotion?

To be able to state this problem more clearly, Korsgaard refers to a more recent discussion as stated by Thomas Nagel. He distinguishes between moral theories as being ‘internalist’ or ‘externalist’. According to an internalist theory, if one has knowledge of a moral judgment, it implies a motive or a reason to act on that judgement. According to the externalist view, such a dependence on knowledge and motivation is not necessary. If something is to be considered a reason claim, a claim about which reasons we have to act, it must according to Korsgaard be capable of motivating. This is what she refers to as the *internalism requirement*⁵⁴.

Korsgaard concludes her argumentation against scepticism about practical reason that it may be due to a misunderstanding of what the internalism requirement requires, namely that we are always motivated by rational considerations. This is too strong a claim; all we need according to Korsgaard is that we are motivated insofar as we are rational. Hence we can allow for some irrationality, and at the same time hold that practical reasoning is instrumental. To be a reason is not the same as always being able to motivate just anyone, all it takes to be a reason is that it can motivate a rational person⁵⁵.

3.3 Reasoning and emotions

⁵³ Korsgaard quoting Hume, Korsgaard 1996b, p. 312.

⁵⁴ Korsgaard 1996b, p. 317.

⁵⁵ Korsgaard 1996b, p. 321.

By claiming that the categorical imperative is different from the Moral Law, I find that Korsgaard paves the way for the possibility of acting in accordance with the categorical imperative, without acting in accordance with the Moral Law. Thus in the case of egoism, where an action may be rational but not necessarily moral, could this be the case? Can the egoist perhaps have the categorical imperative as their acting principle and as such can be rational, and is as such perhaps not completely determined by the laws of nature? To save the egoist from being completely determined by the causalities of nature, I guess requires that the categorical imperative can have a hold on us independently of whether or not there is any moral content. But how can such an imperative motivate us if moral emotions are not on stake? Thus is our reasoning or our reflection, or the rational aspects of such, completely separated from emotions in Korsgaard's view? Or is it possible, or even necessary to see emotions and reasoning as equal aspects, or perhaps perspectives of the same process or procedure?

Kant introduced the emotion of respect as the driving force behind the categorical imperative; the question then arise is this also an option for Korsgaard's view or does she need to find the motivational source of the categorical imperative elsewhere? According to Korsgaard, Kant takes as a starting point that what is unconditionally valuable is the good will, being valuable independent of its consequences. An action is valuable because of the grounds that determine it, not due to its purposes.

Korsgaard explains three different kinds of motivation that Kant advances; to act from duty (from what is the right thing to do), to act from enjoyment (direct inclination) or to act from indirect inclination (as a means to an end). Only to act from duty has moral worth, since duty may be said to be the good will with some restrictions, and so only by being motivated by duty does one act rationally. Hence an act from duty gets its value from maxims "subjective principles of volition"⁵⁶. (Korsgaard 1996b, p. 57) A maxim may be seen as a reason for action, a reason expressing understanding of a principle or law. But how is this unified with the ideas of autonomy and freedom? Korsgaard argues that in deliberation and choice Kant uses the notions of autonomy and freedom as something which we as rational beings attribute to ourselves, as a way to view our reasons or maxims for our actions. When we are engaged in choice for action,

⁵⁶ Korsgaard quoting Kant.

we do not consider ourselves as determined by anything, or forced into it by any desire. Rather, we see ourselves as acting from principles which we adopt freely; it is this voluntary adoption of principles which entails the idea of an obligation. Thus to be obliged does not contrast with our self-perception as free. This does not mean that one cannot take a desire to be a reason to act; it only means that you are not forced to act because of this desire – it is a choice. We are dealing with a hypothetical imperative – a principle telling us what we ought to do to attain an end – in this case, that by following our desire we would come to an end. To be considered a rational maxim to act from, a hypothetical imperative must pass certain tests. It must be a means to your end and “(unless it is morally required) your end must be consistent with your happiness.” (Korsgaard 1996b, p. 14) If Korsgaard agrees in this, I guess she must also agree of the possibility of rational actions that are not moral actions. Rational in the sense of being a maxim which can meet the test of hypothetical or categorical imperatives. In the case of obligation, however, this is not so, that one morally ought to do something expresses necessity to do an action. Moral ought is expressed by a categorical imperative, telling us what to do independently of our private purposes, subjective reasons or desires. But why should this exclude the possibility of a rational ought, an ought not necessarily moral but still expressed by a categorical imperative? But Korsgaard continues supporting Kant’s view that morality is grounded in the human will, and the good will is the source of value.

“There are two ways of being motivated, autonomously and heteronomously. When you are motivated autonomously, you act on a law that you give to yourself; when you act heteronomously, the law is imposed on you by means of a sanction – you are provided with an interest in acting on it.”(Korsgaard 1996b, p. 22)

Thus a will is according to Korsgaard free due to it not being influenced by any external force, including desires. At the same time though, it is causality and hence must have a law. This law cannot be imposed from the outside, and must therefore be chosen by oneself – and it is in this sense we are autonomous. But one concern I have with regard to Korsgaard’s view on autonomy is that this choice appears to be what she claims it is not – causal. Not causal meaning determined by powers outside us, but rather determined from within. Her conception of a self not transcending human nature appears to imply that we are forced by our own nature to make the choice the way we

do. This again it is hard to see how she can combine with her conception of freedom and responsibility. Does this mean that Korsgaard's view on autonomy is rather heteronomous? To this I believe Korsgaard's answer is no. By heteronomous motivation Korsgaard means something optional, since we can chose to change our interests. Autonomous motivation on the other hand is compulsory. But what role do our emotions serve for our motivations?

According to Korsgaard our emotions may serve as an important contributor to our motivations, but then in a heteronomous sense. Korsgaard believes that to act from inclination is like to act based on emotions, or to act instrumentally as a means to an end. Only by being autonomously motivated is an act moral, duty obligates us unconditionally, and only a categorical imperative can manage this. Does this mean that Korsgaard thinks that emotions cannot play a rational or moral role? We have according to Korsgaard, negative moral emotions such as guilt and regret, as a way of punishing ourselves when failing to act from duty. However this punishment or pain is not a threat to our authority over our minds, it only states that since pain is perception of a reason⁵⁷ (both in a backward and forward looking manner), we are capable of experiencing them. Now I will explore one emotion to which Korsgaard pays attention (in a footnote⁵⁸) as possibly having some cognitive aspects – namely respect. The moral emotions of respect are our awareness of the mental activity of a moral reason, an awareness having the character of a feeling. But this means that our emotions come late, and that we have our moral norms, and reasons independently of our feelings. In this view emotions play a role in our rationality, but a very small one. Moral emotion is not the source of moral reason it only perceives it in her view. Respect may be seen as the emotion of taking an interest in and motivate us to take other rational beings into consideration, hence it is natural to assume it plays a role related to the process of the categorical imperative. In this sense it may be said to be rational, and perhaps ensure tat we are not “slaves of passion”. So this may be the solution to the problem concerning how the commands of the categorical imperative motivate us, to why we should follow it. The answer to this as given by the Formula of Humanity is based upon the idea of mutual respect, to consider other rational beings as ends and not as means. Thus the categorical imperative will, according to this interpretation, have a hold on us due to an underlying emotion or

⁵⁷ Korsgaard 1996a, p.151.

⁵⁸ Korsgaard 1996a, p.151.

idea of respect for other human beings⁵⁹. Respect can be viewed as an autonomous motive⁶⁰ as well as an emotion. This however requires that one equates being autonomously motivated with being morally motivated. So by linking respect for the categorical imperative with respect for other people, this could be an argument supporting what appears to be Korsgaard's view - that rational agency always are moral agency. This implies that respect is the only autonomous motive, where autonomy can be understood as a kind of freedom restricted to moral activities. But I don't think this tells the whole story. I see no good reasons why one can't similarly claim that self respect could count as being an autonomous motive. Self respect as meaning respect for the categorical imperative combined with the view that the categorical imperative is not solely a moral imperative but first of all a rational imperative. So the emotion of respect as a motivational force is not in my opinion an argument supporting Korsgaard's and Kant's view that that rational agency is moral agency. I neither find that the emotion of respect can solve the problem of motivation for Korsgaard. Even though she claims to support the Kantian idea of respect as a leading motivational force, her view of emotions belonging to our passive rather than active perspective makes it hard to see how this can be. Rather, Korsgaard interprets Kant in such a way as to make the gap between emotions and our rational capacities larger than would be the case of a 'two world' interpretation. In the 'two world' view, an emotion perceiving a reason would by that very act have rational content. But to Korsgaard active versus passive is the essential distinction, and no emotion whatever it perceives could I guess in her view be anything but passive.

How can Korsgaard claim that because emotions are perceptions of our reasons "then emotions must play an essential role in moral life even on the most rationalistic theory." (Korsgaard 1996a, p. 151) Her view does not seem to allow for motivation to be emotionally driven, but only to have them as its consequence, so I cannot really see which important role she has in mind. Hence the Kantian idea of emotions which Korsgaard appears to support, as a passive, perceptual, perceiving issue leaves them in a different domain to deliberation which is something we actively do. So from that perspective it seems that emotions cannot by nature have rational content, and that she

⁵⁹ Carla Bagnoli argues for a similar interpretation in "Respect and Membership in the Moral Community".

⁶⁰ Carla Bagnoli discusses this issue in "Respect and Membership in the Moral Community".

needs to find other motivational forces than the emotion of respect, or any other emotions, if she wants to avoid Hume's concern that we are slaves of our passions.

Korsgaard mentions desires, inclinations and respect for the moral law as what presents us with our options for what to choose as reasons. So incentives define our domain for what we take an interest in so to consider something as a possible reason could be said as equalling taking an interest in it. Further, Korsgaard argues that "Kant claims that it is impossible for a human being not to be moved at all by incentives; our freedom, rather, is exercised in choosing the order of precedence among the different kinds of incentives to which we are subject." (Korsgaard 1996b, p.165) According to Korsgaard, this is what makes us imperfectly rational beings. If we follow what she refers to as Kant's 'Argument from Spontaneity' the Moral Law would be the first principle of a pure rational will, and the categorical imperative the law of spontaneity. But why should we choose the moral maxim over the maxim of self love? In order to understand this let us first take a look at how Korsgaard understand Kant's distinction:

"The maxim of self-love says something like:

I will do what I desire, and what is morally required if it doesn't interfere with my self-love.

And the moral maxim says something like:

I will do what is morally required, and what I desire if it doesn't interfere with my duty." (Korsgaard 1996b, p. 165)

This seems to be the same distinction as we were dealing with when distinguishing between the egoist and the moral person. But what is it that makes us take an interest in other people why do we find morality important? According to Korsgaard, what we need is an incentive to "identify with the free and rational side of our nature". She continues that this is why Kant made the 'two-world' distinction. This is a distinction situating inclinations in the world of phenomena, making the inclination of self-love a deterministic and naturalistic option. Thus acting from self love is not an act governed by the free will. In this respect, Korsgaard's 'two standpoints' view may be said to make the gap between the person acting from self-love and the moral agent less than Kant. For her to act from self-love I guess would be to act according to a third personal perspective. But if self-love is a naturalistic option why is not also respect so? I believe

this question is difficult for Korsgaard to answer, and I believe the answer has to be that also respect is passive and as such do not belong as an aspect of our moral or rational reasoning's. If she were to claim otherwise I believe she would have to loosen up on her conception of the dualities of reason. Perhaps another way out of it is if her conception of human nature could entail some ideas of virtues.

One intuitively appealing idea I find, is that motivation for action, also in a rational sense, is intimately connected with the emotion of taking an interest in something. Korsgaard acclaims Kant's view that all actions must have an end, and that each choice is determined by such an end. To adapt a moral end is, in this view, a virtue. There are several moral ends and hence several virtues all in accordance with the Supreme Principle of the Doctrine of Virtue:

“Act according to a maxim whose ends are such that there can be a universal law that everyone have these ends”⁶¹ (Korsgaard 1996b, p. 178)

This principle allows for what she calls internal freedom, freedom in the form of the virtue of internalizing moral ends. All virtues represent a duty, an obligation to act according to the Moral Law. But in what sense do we then talk of freedom if it means obeying a law? To answer this Korsgaard quotes Kant as saying that we are free to construct our own ends in the sense that no one else can force an end upon us; it is entirely up to us which ends we adopt – and this is what Kant means by internal freedom. When adopting an end, given that it is a moral end, we take pleasure in it. But that we take pleasure in adopting an end does not mean that achieving pleasure motivated us, since only to act from duty has moral worth and can move us to act in a rational way. Thus Korsgaard distinguishes between the emotions which make us adopt a purpose and those resulting from adopting a purpose⁶². To act rationally means to act independently of our own inclinations, this means not to act according to our own interests like those which are emotionally based. But how can Korsgaard claim that she wants to make Kant's harsh division of our faculties smaller, when she in many respects appears to do the opposite? One way out of this for Korsgaard could be to push Kant's theory of virtue in an Aristotelian direction and by that attribute to emotions some

⁶¹ Korsgaard quoting Kant.

⁶² Korsgaard 1996b, p. 59.

rational impact. There is a new interest in Kant's view on virtues. Nancy Sherman writes in the book *Making a Necessity of Virtue, Aristotle and Kant on Virtue* that in

“The Doctrine of Virtue, Kant recognizes the duty to develop emotions as part of our duties of virtue.... our end-setting capacities (i.e., rational agency) are sustained and developed” (Sherman 1997, p. 125)

Hence her point is that emotions as part of our human nature will have an impact on morality due to it being grounded in human nature. Therefore it is our duty to develop these emotions in order to improve our rational agency and end setting capabilities. These ideas go well with Korsgaard's ideas of human agency as the source of normativity, and her more practical approach to rational agency than the ideal view Kant supports. However I cannot see that it harmonizes with the Kantian idea of rational nature.

“Practical reason is a faculty of ends, so if there is pure practical reason there must be necessary ends. This means that there are duties to have these ends, duties of virtue.”⁶³(Korsgaard 1996b, p. 20)

Humanity is the end set by reason, but to accomplish virtue these ends must be achieved freely. How successful one is concerning these virtues depends upon whether one manages to carry out in the duties of virtue from a moral motive or not. We are anyway obliged to try, even though perfection is not realistic she concludes.

3.4 Naturalism and the nature of the self

Korsgaard's conception of identity is to me unclear. She operates with the idea of a practical identity, which is an identity we construct – an identity emerging through action or reflection. But is this the only conception she has of a self? When describing the procedure of reflective endorsement, it is a procedure of distancing ourselves from our inclinations. She says that there is an 'I' as if it is over and above ourselves, an 'I' making it possible for me to reflect upon my own thoughts. How is this 'I' to be understood? What she refers to I understand to be that the 'I' is our self-consciousness,

⁶³ Korsgaard quoting Kant.

our awareness of our own mental activities. If so, what does it imply – that she supports some kind of naturalism?

In “On Naturalizing Kant’s Transcendental Psychology”⁶⁴ Allison argues against interpretations of Kant naturalizing his conception of ‘a self’. With ‘naturalizing’ Allison appears to mean something similar to having a causal or functionalistic account of representations of objects, giving rise to an understanding of the self as being empirical. Empirical in the sense, that a synthesis is understood as a conscious act combining elements from one cognitive state to create a new one⁶⁵. Thus cognitive mental states that are connected synthetically are so in a causal manner. The ‘transcendental unity of apperception’ is, in this view, exactly such a synthesis giving rise to new cognition. This does sound as being close to how Korsgaard may be read. This is in contrast to the view which I assume Allison holds that this unity of apperception is the starting point of the transcendental deduction. He interprets Kant as with identity or unity meaning a necessary condition for cognition, an identity we think a priori, and not something which one is aware of or identifies oneself with – rather it is fundamental for us being aware.

“In short, conceptual recognition (which is what the objectivating synthesis amounts to for Kant) is an inherently reflexive and, therefore, self-conscious act. I cannot perform it without an awareness of what I am doing (a consciousness of the act), although I can certainly do it without reflecting that I (Henry Allison) am doing it.”(Allison 1996, p. 62)

This idea of an ‘I’, is based on the idea of the self as not being an object to which one can have cognition either in a sensible or an intellectual way. This is in contrast with how I understand Korsgaard’s view of identity, as something emerging exactly through deliberation. It is by distancing oneself from one’s inclinations, by reflection, meaning seeing central aspects of oneself from a distance one comes to have an identity. Hence the naturalistic view as given here seems to have more in common with how I understand Korsgaard’s view than Allison’s Kant-based view. I will now see in what sense she herself may consider her view to be naturalistic.

⁶⁴ Allison, *Idealism and Freedom*, pp. 53-66.

⁶⁵ From Kitchers formal def. of a synthesis, Allison, *Idealism and freedom*, p. 54.

“In one sense, the account of obligation which I have given in these lectures is naturalistic. It grounds normativity in certain natural – that is, psychological and biological – facts. I have traced the normativity of obligation to the fact of reflective consciousness and the apparent normativity of pain to the fact of simple consciousness, together with the nature of an animal. My account does not depend on the existence of supernatural beings or non-natural facts, and it is consistent with although not part of the Scientific World View. In that sense, it is a form of naturalism. But in another sense it is not. In another sense, a naturalistic view identifies normative truth with factual truth.”(Korsgaard 1996a, p. 160)

It is very difficult to read this in any way than that she grounds normativity in the empirical, in psychology and biology. But from where then could necessity of obligation arise, if not from the laws of nature? To this she claims that it only needs to be consistent with the scientific world view, it is not a part of it. So what is it a part of? At least the causalities of reflection must be of a very strong kind, as though being laws of nature. In several places in her text she points to passages in Kant saying something like this, so I presume this is how she likes to interpret his view. It is a view focusing more on human nature than on rational nature, and in this sense she seems closer to Hume than Kant in this respect. However, a problem with this view of reflection is that it is difficult to argue that our thinking is not determined. Therefore, are there any other options than an ‘I’ understood as a ‘transcendental unity of apperception’ or an ‘I’ threatened by naturalism? If so, is this a conception of an ‘I’ which is available to Korsgaard. I will now explore these matters.

Korsgaard’s view of the first person perspective rests upon a division between our incentives and on our more rational capabilities. It is by distancing oneself from impulses such as our emotions that we are capable of making rational decisions. Thus deliberation or first person perspective is in this sense somewhat dualistic or divided. What is commonly characterized as the standard view of rationality holds that rationality can be reduced to the notions of desires and beliefs. This means that acting rationally implies acting in accordance with certain beliefs and desires. Korsgaard has a different view holding that desires play a different role. The very fact that you don’t just wait to see which desire becomes a reason, you actively choose between your conflicting desires by having reasons for or against them. These reasons and not your desires are expressions of your will, Korsgaard explains. (Korsgaard 1996, p. 370) Thus to act on a desire requires that one finds oneself to have a reason to do so - rationality is

about responding to reasons. However, for Korsgaard it does not end here, which reasons we have is a matter of being in coherence with the principles of reason.

One who also opposes the standard view of rationality is Scanlon. But to him the notion of being a reason is primitive. Desires in the standard interpretation according to Scanlon, are psychological states serving both as motivating actions and as initiating justifications. However, Scanlon disagrees in giving desires this crucial role, arguing that the notion of being a reason better serves these purposes. This is not the same as saying that recognizing something as a reason motivates us. Scanlon disagrees in the very idea held by the standard conception that we should have some sort of motivational state of mind in addition to the considerations we make concerning which reasons we have. If someone is rational and finds something to be a reason to act, it explains that person's intentions and actions according to those reasons. However this is not the same as saying that a person is motivated by recognizing something as a reason to act. According to Scanlon we do not need some external (to our considerations) motivational state, it is sufficient with the state of mind we are in (our attitude) when making the consideration that something is a reason. Hence our desires have a cognitive aspect when being involved in seeing something as a reason. These attitudes are in a sense desires; the crucial point being that these desires do, not imply an additional mental state different from the state from which we make our considerations. 'the language of reasons, as opposed to mere desires, is crucial to an adequate description of the structure of our own practical reasoning and also to our relations with others'.(Scanlon 2000, p. 77). Scanlon links the concepts of desires and the structural features of our reasoning together, rather than seeing them as separate issues. He argues for a strategy or method for evaluating our judgements about which reasons we have. In the moral case deciding what to do is to see which reasons one cannot reasonably reject. One of his main arguments for this view is that in dealing with moral issues one is more concerned about reasonableness rationality. It can also be argued to provide a more realistic and practical procedure in decision making, than the seemingly hopeless project of identifying all of one's reasons and then deciding which one is the best. Thus in a sense he presents an account which is more practical than that of Korsgaard. More importantly though he has a view of deliberation that does not rest on the duality that Korsgaard's view does. I find that one central idea that Alison and Scanlon seems to

share is the idea of first person perspective as something one is in rather than something one is being self-conscious about. This is, I suppose, is what one more frequently considers as a first person perspective. Korsgaard's first person perspective is divided, and as such gives rise to several problems in my opinion.

3.5 Conclusion

I have argued that I am not convinced that Korsgaard achieves her aim of softening Kant's dualisms. Her view of the process of deliberation I find problematic. To step back from one's inclinations means to put behind all our practical identities except the one we all have in common - that we are human beings. If this is the case for everyone deliberating, what then is the difference in doing this and from taking the third person perspective of oneself – or as I guess Korsgaard would put it – the perspective of another self? If in deliberation one takes away every emotion or inclination, everything which is peculiar to that person, what then is the difference between another self and a third person? I cannot see that Korsgaard gives an answer to this. However, if we then accept that deliberating is to step back and reflect about our inclinations *as if* we were someone else, does this not then make it very close to the theoretical perspective of explanation? Then her two standpoints' appears to collapse into one.

I find that the main problem is due to Korsgaard's demand to our reasoning as being causal in the sense of as if being a part of nature. This demand is unnecessary strong and makes the dualities of reason harsher than for Kant rather than the opposite. And it makes one question why should the fact that we are the source of the law which binds us necessarily lead to freedom and responsibility? The conception Korsgaard presents of 'the self' does not obviously provide a support to this view. Then the less introspective conceptions of 'the self' as provided by Scanlon and Allison, for very different reasons, may be better candidates to meet the quest for responsibility. But perhaps doesn't the interpretation of Korsgaard as defending some sort of naturalism tell the whole or true story? In the next chapter I will explore if the necessity she associates to our moral obligations can be justified by transcendental arguments instead of resting on causality.

4.0 Autonomy and freedom

Why should the fact that we are the source of the law which binds us necessarily lead to freedom and responsibility? Korsgaard's arguments give rise to confusion of whether autonomous agency equals moral agency, rational agency, both or neither. In this chapter I will try to explore these concerns further.

Korsgaard's division of the categorical imperative from the Moral Law appears to create problems of justifying both the categorical imperative and the Moral Law. The categorical imperative, as constructively justified, gives rise to self-referential problems, while the force of moral obligations can be questioned as well. Perhaps the problem is due to different expectations about what a law is? So, what is a law? For Korsgaard, to answer this question is vital to understanding similarities and differences between her and Kant's position. One reason for this is because they both argue that the categorical imperative requires that what we accept as a reason to act, must take the form of a law. To count as a law in their sense, is both owing to the judgements' universality and to their necessity⁶⁶. However, as I have argued, their idea of what is meant by universal is not obviously the same. The difference is possibly due to how they argue and, in particular, due to in whether their arguments are transcendental or not. On the other hand their conception of necessity seems to be rather similar, at least in claiming to have the same force. This, I presume, is due to their apparent common conception of the relationship between freedom and law, and in how this conception is supposed to balance the determinism they associate with the world of nature. In this chapter I will explore further these possible differences and similarities. One important question is in what sense, or if at all, Korsgaard's arguments from the categorical imperative to the Moral Law can be said to be transcendental. If so, then perhaps the necessity of moral obligations is secured, as well as the idea of our moral identity as inescapable?

4.1 Transcendental argumentation

So what is meant by transcendental argumentation? A transcendental argument is commonly understood as an a priori argument, an argument independent of experience. Synthetic a priori judgements or transcendental judgements have their status as a priori

⁶⁶ Theo van Willigenburg argues in a similar manner in 'Shareability and Actual Sharing: Korsgaard's Position on the Publicity of Reason' p. 174.

by necessity because they have their source in the ‘transcendental unity of apperception’, a kind of ego which serves as a necessary condition for our self-consciousness. It is a condition of knowledge, but since it is not an object of knowledge we cannot know anything about it. If one interprets this ego as having universal force, it could be seen as what limits our freedom to also concern others. Kant argues, according to Allison,⁶⁷ that practical freedom and transcendental freedom are dependent issues, where we do not have the latter without having the former. Transcendental freedom is what Kant's Third Antinomy concerns, based upon our ability to initiate a state entirely from ourselves – or spontaneously. Transcendental freedom is what Kant denotes as a ‘pure transcendental Idea’ or concepts of reason, meaning something independent of any object given in experience. Transcendental freedom is contrasted with mechanical causality and determinism underlying any kind of time order. Practical freedom, on the other hand, concerns human agency and our ability to recognize something as being a reason, a freedom depending upon the pure transcendental Idea. Transcendental arguments are commonly traced back to Kant⁶⁸ as an argument against epistemic scepticism. The notion of a ‘transcendental argument’ is not commonly understood as a proof, but rather as a looser form of justification. One of the things Kant wanted to demonstrate by a transcendental argument was that we cannot have experience without the existence of time and space. This was meant as an argument to the sceptic of the existence of things outside us. In general, a transcendental argument has the following structure⁶⁹: one wants to show that a judgement is true, and the task is to find other judgements which are necessary conditions for this judgement to be true. Transcendental arguments are commonly used to argue against scepticism by showing that scepticism is incoherent or inconsistent. This is done by depending upon unquestionable, unavoidable conceptions of thoughts, experience or language. Commonly such an argument takes as a starting point some presuppositions or concepts which are necessary in order for the sceptic to pose their challenge, and then goes on to show that these very presuppositions dismiss the challenge. A modern example of a transcendental argument is given by Hilary Putnam, arguing that we are not ‘brains-in-

⁶⁷ Henry E Allison ‘Kant’s transcendental idealism’ p. 315.

⁶⁸ Bardon, A. 2006, “Transcendental Arguments”, *The Internet Encyclopedia of Philosophy*, ‘, pp. 1-2, www.iep.utm.edu/t/trans-ar.htm

⁶⁹ Filosofi Leksikon , Red. Poul Lübcke, p. 434.

a-vat'. This he does by arguing that if someone asks the question 'am I a brain-in-a-vat?' either the answer is no or the question is meaningless⁷⁰.

Kant made several transcendental deductions, among them one of the Moral Law. This deduction was based upon the assumption that the categorical imperative is the Moral Law, so this deduction is intimately connected with the foundation of the categorical imperative. But, if Korsgaard cannot be said to argue in a transcendental manner, is that a problem for her position? I suppose the answer to this is both yes and no. 'Yes' owing to her strong Kantian interpretation of law and its necessity and from where her conception of necessity is to come from if not established by a transcendental argument, the alternative appears to be naturalism. 'No' because of common criticisms of transcendental arguments, since what is commonly referred to as Kant's 'transcendental psychology' is among what's most frequently criticized in his theoretical philosophy⁷¹. This criticism is often due to dismissing what is taken to be Kant's view of mental activities as the "imaginary subject of transcendental psychology"⁷². One influential critic is Barry Stroud, who in his paper "Transcendental Arguments"⁷³, presents a case against transcendental arguments – at least in the powerful sense of dealing with 'outer world' scepticism. He claims that there is no transcendental proof against outer world scepticism, and argues for a more modest argument claiming that it is enough that we believe in an external world. Stroud argument is⁷⁴ loosely that there are statements which belong to what he calls a "privileged class". That is statements which cannot avoid being true when uttered by someone. Even though there are such statements, the sceptic can always argue that there are not so. If the task of transcendental arguments to the sceptic is what Stroud claims, to show that what the sceptic doubts is in this privileged class, then transcendental arguments are shown not to work. Korsgaard does not seem to be, at least not in an explicit manner, occupied by the very central aspects of Kant's philosophy dealing with issues such as 'transcendental freedom', 'transcendental deduction', 'transcendental Ego' and so on. Thus, there are good reasons to suspect that her ideas of freedom and

⁷⁰ Skidmore 2001, p. 122.

⁷¹ Most influential among the critics are said to be P. F. Strawson in his book *The Bounds of Sense*, according to Allison, *Idealism and freedom*, p. 53.

⁷² P.F.Strawson, *The Bounds of Sense*, London: Methuen & Co. Ltd, pp. 31-32.

⁷³ Stroud Transcendental arguments

⁷⁴ As explained by Skidmore 'Scepticism about practical reason: transcendental arguments and their limits' and by Stroud in his 'Transcendental Arguments'.

identity do not rest upon transcendental arguments in this sense. However in one place in her text, she claims that her argument for the Moral Law is transcendental. The question then arises in what way her argument is transcendental, and how can Korsgaard's argument deal with the kind of criticism as directed by Stroud. To argue from the categorical imperative to the Moral Law – or at least a step on the way⁷⁵, is where Korsgaard claims to use a transcendental argument. I will also see if her conception of 'transcendental' as understood from this argument can explain the claimed force normativity is supposed to have on us, and the necessity of moral obligations. Or is her theory in need of another undetermined starting point for autonomous activities than that represented by Kant's idea of spontaneity, another starting point securing our freedom from the causality of nature? But the question is if this is possible. According to Allison, Kant would claim that all theories except his own are heteronomous. Since these theories have the form

“‘I ought to do something *because I will something else*’ (Gr 4:441; 111). Moreover, since this interest, whether sensuous (e.g., in one's own happiness) or rational (e.g., in ontological perfection) reflects the “natural constitution” of the subject, Kant once again asserts that in all such cases it is “strictly speaking...nature which would make the law”(Gr 4:444; 112)” (Allison 1990, p. 100)

So I guess one concern is whether Korsgaard's theory has this form, something which I find to be a plausible interpretation is exactly the case.

4.2 Transcendental argument for the value of humanity

Whether or not Korsgaard's arguments from the categorical to the Moral Law are transcendental or not is controversial, and several proposals have been made. One of the problems is due to disagreement concerning what is meant by a transcendental argument, another problem is due to different readings of Korsgaard. Kant's transcendental deduction of the Moral Law is summarized by Korsgaard as an argument which “connects freedom and reason through the capacity of reason for pure spontaneous activity which is exhibited in its production of ideas.”(Korsgaard 1996 b, p. 161)

⁷⁵ Stern 'The Value of Humanity: Reflections on Korsgaard's Transcendental Argument' p. 15

As we have already discussed, this is different from how Korsgaard argues in order to establish the Moral Law. Therefore let us take a look at Korsgaard's arguments.

“The argument I have just given is a transcendental argument. I might bring that out more clearly by putting it this way: rational action exists, so we know it is possible. How is it possible? And then by the course of reflections in which we have just engaged, I show you that rational action is possible only if human beings find their own humanity to be valuable. But rational action is possible, and we are the human beings in question. Therefore we find ourselves to be valuable. Therefore we are valuable. (Korsgaard 1996a, pp. 123-5)

With practical normative scepticism⁷⁶ Korsgaard refers to doubts about whether rational actions exist or not. Thus her argument can briefly be described as taking as a premise which also the sceptic would also accept; the existence of human agency and that in order to act we need reasons to act. Here I suppose that we are speaking of instrumental reasons and not moral reasons. For her, the unconditional starting point is owing to our practical identity of humanity, the identity which is a necessary identity to have in order to have any practical identities at all. It is the identity based upon us as being conscious reflective beings, capable of having and requiring reasons for our actions. Thus rational human agency is possible which for Korsgaard includes the idea of a free will. The very characteristic of human agency is our reflective consciousness and our search for reasons, and that these reasons are in coherence with our practical identities – and, in particular, the identity of humanity. Hence humanity is our unconditioned practical identity, the identity which stops the regress. This is not possible if you do not value your humanity.

However, her conception of value is not one demanding a strong transcendental argument. Value in her opinion is first-personal, something which we find ourselves to be, rather than something evaluated from the outside. Therefore we do not need to relate to Hume's scepticism concerning whether there is a world outside us, to whether there is anything out there of any value. Hence she does not need to give a strong transcendental argument; her argument can be characterized as being modest⁷⁷. I suppose this is so due to her dismissal of what she refers to as substantial realism. Her idea of procedural realism would make the appropriate scepticism to be whether or not

⁷⁶ Korsgaard 1996a, p. 164.

⁷⁷ Skidmore 2002

our procedures or reflections actually exist, which may seem to be obvious – unless of course we actually are ‘brains-in-a-vat’ and they would only exist from a third person perspective which I guess imply a more substantial realism.

But how can Korsgaard’s modest transcendental argument establish any kind of necessity? It is hard to see how her arguments ending with the Moral Law, can ensure obligation and morality, the rationality and necessity at which Korsgaard aims at. In what sense is Korsgaard’s argument transcendental? Is it only due to its form as described in the introduction, to the structure of what is commonly considered a transcendental argument, or does it have anything to do with the necessity it is supposed to provide? To explore this concern I will examine James Skidmore’s formulation of the argument⁷⁸ in which the aspect of necessity is more accentuated. He explains Korsgaard’s argument as working back from the premise that our humanity is due to that we act for reasons, to its necessary conditions. These necessary conditions he explains as being the “the necessity of taking something to be normative, of developing a ‘normative identity’”, and further the necessity of valuing one’s humanity or agency as an end. This, Skidmore claims, is a transcendental argument because it provides ‘one rationally necessary end: our own humanity’ it is, as such, an argument going further than instrumental reason. Thus to value one’s humanity is not something one arrives at by means ends reasoning, it is an end by necessity owing to our rational powers. But, if these powers are questionable, that which we are dealing with is human nature or some kind of causality. Would that in his view qualify as a basis for a transcendental argument? Probably not, I believe to accept this as a transcendental argument requires a rather Kantian reading of Korsgaard.

4.3 Universalizability and the necessity to see the value in others

This first transcendental argument wishes to establish that scepticism of reason is a mistake. However, to claim that what is ultimately valuable is humanity, does not rule out egoism. To accomplish this Korsgaard advances one further argument, an argument against scepticism of moral reason. My question is then: is this also a transcendental argument? As we recall she parallels her argument with Wittgenstein’s private language argument. This is an argument which is commonly understood as a transcendental

⁷⁸ Skidmore 2002, p.134.

argument in the strong sense⁷⁹. But is this how Korsgaard reads him? And if so, does this imply that Korsgaard's argument too is transcendental in this strong sense? No, it does not; Skidmore argues⁸⁰ and refers to Korsgaard's interpretation of Wittgenstein as saying:

“She suggests that his argument, beginning with the fact that linguistic meaning is normative, establishes the publicity of language as a necessary condition for this normativity.”(Skidmore 2002, p. 135)

This due to that we cannot have a language which is in principle incommunicable. And similarly she wants to do the same kind of argument against private reasons. Since reasons must be normative in order to be reasons they cannot be egoist or private reasons, but must be public. By valuing our own agency as such, we value the agency of others. We have the same public reasons which for that reason also must be moral reasons. Skidmore claims that this argument might be transcendental in a modest sense, and as such escape Stroud's criticism. But that is not the same as saying that the argument is sound. He claims that the agent-relative conception of egoism is not the kind of privacy we need in order to parallel the private language argument, Agent-relative reasons are not incommunicable in principle to anyone. Skidmore brings this argument one step further:

“transcendental arguments for *moral* reasons would seem to be doomed to fail due precisely to the *intelligibility* of the rejection of such reasons. Whether or not the egoist's account of reasons is ultimately adequate, it certainly appears to be a *coherent* account. After all, there is nothing incoherent or inconsistent in imagining someone who reasons exclusively from the point-of-view of prudence, however impoverished such a point-of-view might ultimately prove. As long as this is true, transcendental arguments will not succeed in defeating the egoist. They cannot demonstrate the unintelligibility of what, in the end, is perfectly intelligible.”(Skidmore 2002, p. 138)

One who I interpret as disagreeing in this is Theo van Willigenburg. He uses as we recall the terms shareability and actual sharing to make the same distinctions. If we are to act according to the categorical imperative, the reason which we choose has to be law like. According to him this means that it must be universal – which he equates with

⁷⁹ Skidmore 2002, p. 135

⁸⁰ Skidmore 2002,

shareable, and it must be a reason by necessity. So he argues that we cannot but share what is shareable, and that Korsgaard gives a transcendental argument for this claim.

Whether Korsgaard's arguments from the categorical imperative to the Moral Law are transcendental or not is hence questionable. But either way there seems to be a problem by arguing away from the egoist. I cannot see that Korsgaard manages to accomplish this.

Korsgaard appears to run into two problems. First, how can her constructivist view ensure a priori the necessity she seeks for the categorical imperative. Further, how can this necessity be also captured in the moral case? But why would what commands the free will need to be universal in the first place if the universalizability criterion according to Korsgaard cannot get us to morality? Korsgaard's reply⁸¹ to this I understand as being due to the very idea of obligation, that normativity requires some sort of law likeness or regularity. It is this law-like formulation universalizability is meant to capture. It is meant to capture the necessity which the imperative is meant to have on us, the power of causes. Korsgaard asks why these ideas of power and normativity cannot stand alone? In answering this, she refers to both Hume and Kant as agreeing that causality and regularity are dependent issues. She uses the relation between cause and effect and the necessitation binding them, as an analogue to the relation between agent and action.

“I need to will universally in order to see my action as something which *I do*.”(Korsgaard 1996a, p. 229)

Regularity or a normative principle of the will is to ensure some sort of unity or existence of agency consisting over time. Hence there is an a priori universal principle empowering our decisions making us into agents persisting in time, she argues. But in what sense is it a priori, given by a transcendental argument? If she is right I suppose what this means is that our egoistic decisions have some normative force a priori grounded. But does this mean that our rational decisions are determined in some sense? No, Korsgaard replies “Laws which cannot be violated cannot be followed either” (Korsgaard 1996a, pp. 231-232) When you give yourself a law as an act of the will. She continues that essential for an act of the will I that you must make a claim to

⁸¹ Korsgaard 1996a, p. 226.

universality, it must be valid on other occasions and in this sense be general. Is this the same that as that an act of the will in Korsgaard's opinion is primarily linked to rational decisions rather than to moral decisions, and that moral decisions are a subset of rational decisions? In that case this would oppose her view of moral normativity as superior, and our moral identity as inescapable. Is that consistency with one's own reasons over time are more important than consistency with reasons which we share?

If one as a thought experiment accepts that her arguments from the categorical imperative to the Moral Law are transcendental in the modest sense. What does it mean? If Korsgaard's transcendental arguments are modest ones, so Stroud's criticism does not apply to her, does this mean that she has managed to equate rational agency with moral agency? She does not give a 'substantial' transcendental argument, but argues for something weaker which does not need to assume the existence of an outer world, of rational agency or of moral agency. The question then is, does she accomplish to argue for what she has set forth? I don't think so. I agree with Skidmore when claiming that transcendental arguments "despite Korsgaard's efforts, they cannot succeed in establishing a rational foundation for morality." (Skidmore 2002, p. 121) That Korsgaard seems, in general, to prefer modest transcendental argumentation, does not that imply that she has to give up the idea of synthetic a priori, the idea of objectivity and of necessity? To Kant transcendental deductions was the method to ground certain concepts and principles as objectively valid. Is this an option for Korsgaard? It is unclear in what way Korsgaard's arguments are transcendental other than purely structural, following a procedure similar to what is commonly known as transcendental argumentation. But I suppose there is one more aspect to this way of reasoning, namely that transcendental refers to a foundation outside of our experience. So how does she ground her theory, where does necessity come from? It may be that what she has in mind is a procedure, but not in the sense I have just suggested.

"the transcendental distinction is not *primarily* between two kinds of entity, appearances and things in themselves, but rather between two distinct ways in which the objects of human experience may be "considered" in philosophical reflection, namely, as they appear and as they are in themselves." (Allison 1990, p. 4)

Perhaps this can be expanded to also relate to Korsgaard's view, that the transcendental distinction is about two different perspectives that one may take in philosophical

reflection? In some places she suggests what can support this interpretation, saying that objectivity is due to the way we reflect. If so, objectivity is procedural. But again, I cannot see how she can avoid serious problems of self-reference when doing this. Korsgaard's transcendental argument was, as we showed, not powerful enough to work as epistemic justification. It only took us as far as to what we can believe. Therefore, how does she move on in order to justify these beliefs? A modest transcendental argument may not be an argument capable of proving that external-world (in Korsgaard's case the existence of rational and moral agency) scepticism is wrong, but to aims for the less ambitious task that we are to believe in such an outer world or in rational action. The problem then is that even if we succeed in establishing this, the sceptic may not be convinced since something we believe may, after all, turn out to be wrong. Therefore, even though we believe something - even by necessity - in addition we need to justify⁸² this belief. I presume this is what Korsgaard means in her appeal:

“If you think reasons and values are unreal, go and make a choice, and you will change your mind.”(Korsgaard 1996a, p. 125)

Justification is due to conviction. Hence perhaps this conviction is the answer to the question on how the categorical imperative can have a hold on us even though it is different from the Moral Law? But then what Korsgaard seems to be doing is to justify one belief – the belief in the existence of rational agency or in practical reason – with another belief – the conviction in the existence of a free will. This associates more with Quine's holistic version of reflective equilibrium than a Kantian theory. Perhaps that is not a problem for her to give up, since by leaving the transcendental in a strong sense behind, Kant's way of grounding his system is not available to her anyway. For how can spontaneity from the self be anything other than naturalistic, if the self it comes from is that of human nature and not a more universal source, universal in the Kantian sense? It seems as though Korsgaard's avoidance of the transcendental and her orientation towards Hume leaves her with a notion of universality not capable of grounding the categorical imperative in anything other than something private or in nature. Thus she ends up with the problem of self-reference, defining the very principles of mental activities by a procedure of the very same mental activities. But could this conclusion be

⁸² Arguments in accordance with Skidmore 2002, p. 124.

due to a misunderstanding of Korsgaard, and that she with human nature means something more in accordance with the Kantian idea of a rational nature? In order to shed light on this, I will explore some ideas from another interpreter of Kant – Henry E. Allison.

4.4 Spontaneity, autonomy and authorization

Allison's interpretation of the deduction of freedom or of the Moral Law is based upon his conception of *transcendental idealism* – which we recall as the two aspect account. Henry Allison argues for an interpretation of Kant emphasizing the connection between spontaneity and apperception as being intimate and central, where by apperception he means our awareness of 'I think'. He quotes Kant as saying that apperception is "an act of spontaneity". My concern is whether or not Korsgaard adopts these kinds of ideas or not. This is important in understanding how she distinguishes between rational nature and human nature. Central to Allison's discussion of freedom is what he considers two varieties of freedom: autonomy and spontaneity. Spontaneity he claims to be rational agency's ability to be self-determinant on the basis of general principles. (Not only moral principles). Autonomy, on the other hand, is practical freedom or moral agency's ability for self-determination. Further, according to Allison, in the case of a rational deliberator, spontaneity is:

"the capacity to determine oneself to act on the basis of objective (intersubjectively valid) rational norms and, in light of these norms, to take (or reject) inclinations or desires as sufficient reasons for action."(Allison 1990, p. 5)

Hence this makes a rational decision dependent on something additional to the causality of one's psychology or mental state it also rests on the act of spontaneity. Korsgaard appears to neglect this aspect of Kant's argument so the question remains: who is the author of the law that governs us? What is really Korsgaard's conception of this? In "Realism and Constructivism in Moral Philosophy" Korsgaard outlines Hobbes's idea of a sovereign: Hobbes said that morality is the solution to the problem of what the situation would be like without morality. As a horror example he describes a state where everyone is at war with one another. In order to solve this situation he argues that people must be motivated to engage in some sort of contract with each other. This

contract would be to give up one's freedom to a sovereign capable of reinforcing moral and legal law and order followed by everyone. In "Reason, Humanity, and the Moral Law" G.A. Cohen reformulates some dilemmas originally stated by Hobbes concerning the source of the law which binds us:

"You might think that, if you make a law, then that law binds you, because you made it. For, if you will the law, then how can you deny that it binds you, without contradicting your own will? But you might also think the opposite. You might think that, if you are the author of the law, then it cannot bind you. For how can it have authority over you when you have authority over it? How can it bind you when you, the lawmaker, can change it, at will, whenever you like?"(Korsgaard, 1996a, p. 167)

Cohen describes the main difference between Kant and Korsgaard as concerning who is the author of the law that makes claims on us. For Korsgaard it is the subject by constructing one's practical identity who indirectly is the author of this law. For Kant, on the other hand, Cohen argues, the situation is more equivocal⁸³. The subject is the author of the law in the sense that the law comes from the will. This is necessary since to only obey laws from the outside is slavery, heteronymous as such. However, if the law only came from my own will, how can it make a claim on me? To this Cohen answers that Kant's individual is designed by nature to make the law by the authority of reason. For Kant, the sovereign is our reason, and the source of the imperatives of morals cannot be that of human nature or human reason. Only by a transcendental grounding of obligation a priori having their source in pure reason can one avoid Hobbes's concern about the sovereign. To Korsgaard this solution may not be available.

"For Korsgaard, morality is grounded in human nature, and that difference between her and Kant is consequential here, for Kant has a ready answer to Hobbes's argument about the sovereign, whereas Korsgaard may have no answer to it, because she has abandoned the element of Kant that transcends merely human nature"(Korsgaard 1996 a), pp. 172-174)

By claiming that the categorical imperative is different from the Moral Law she in some way moves closer to determinism. To solely base the lawmaking on human nature, and if one wishes to avoid Hobbes's concern that if you have authority over the law how can it have authority over you? The only way for Korsgaard to avoid this, if insisting that

⁸³ Korsgaard 1996a, p. 171.

the categorical imperative is not the Moral Law, is to place authority on human nature, but then we are no longer free but rather programmed by something/someone other than oneself.

Kant by linking the idea of a categorical imperative with the idea of a Kingdom of Ends can be interpreted as been giving the categorical imperative content. It could also be seen as opening up the way for giving emotions such as respect and also experience and knowledge some impact on our practical deliberations. By her concept of Practical Identity, Korsgaard may appear to be doing the same. However, as I understand her, only our reflective self-consciousness - the Practical Identity of being a human - is capable of obligating us. Thus by arguing that the categorical imperative in some sense is only a law, that it is pure structure, and at the same time having human nature as its source, is that not the same as putting form back to nature? If determinism comes from within - from oneself - but a self only as representing a kind or a species, does it make it any better than if we were forced by some outer powers? If reflection is not something which we can escape, and its results are due to laws, then the interesting question I do not find to be whether the principles or laws forcing our reasoning is as if they were laws of nature, or if they actually are the laws of nature. In any case, in my opinion it is a view putting too strong a claim on the causality of our reasoning - namely the claim of it being compulsory.

4.5 The value of choice and the causal thesis

To Korsgaard autonomy seems to mean something else than what Kant had in mind, and in many respects one can argue that her conception of autonomy is heteronomous. But does this not imply that the categorical imperative is not categorical, but hypothetical? If it is hypothetical, is that necessarily a problem? The problem can be solved if one does not require from the idea of freedom that it somehow must entail law meeting the strong understanding of causality. I believe that the very need to connect freedom to law (meaning by necessity) is due to be able to deal with what is considered to be a deterministic mechanical world.

If one's conception of the scientific world was not so rigid, perhaps one could have a conception of freedom and choice not so much in need of strong causality and law in our thinking and reasoning either. Scanlon presents a different view, offering an

account of right and wrong which is heteronomous and, at the same time, arguing for a categorical imperative. But how can he make claims to both a categorical imperative and heteronomy? As we recall Scanlon does not make the distinction that Korsgaard does in that the categorical imperative is different from the moral law, rather he promotes an understanding of universal as meaning 'for everybody'. He gives an account where the moral authorities which Korsgaard claims to be grounded in rational or human agency, in Scanlon's view has its foundation in 'other aspects of our lives and our relations with others'. (Scanlon 2000, p. 6) Scanlon offers a view of responsibility not based upon the view of autonomy but rather based upon principles given external to us.

To Korsgaard, value is as we recall something we impose on the world, it is of our creation. But her view, I have argued, has a problem of being self-referential; at least it is unclear what kind of foundation she gives to prevent this. I turn now to explore another view on value which I find does not run into the problems that Korsgaard's view may do. Scanlon argues against the common view that 'well being' is what he calls a 'Master value'. In this view to be valuable or good, is a state of affair to be promoted in order to bring well being to the individual involved. Scanlon argues that 'good' is not the only value to promote, we also have values separate from 'the good' such as friendship, excellence in art and science, love and the value of human life in itself. His account is two-fold; first, his idea that value is not only a teleological issue, and second he considers the notion of value in an abstract manner:

"being valuable is not a property that provides us with reasons. Rather, to call something valuable is to say that it has other properties that provide reasons for behaving in certain ways."(Scanlon 2000, p. 96)

This is what he refers to as the 'buck-passing' account of value and goodness owing to that to be good or valuable is not to be considered as a reason which to respond in themselves. What we respond to, what serves as reasons for us to act is that to be good or valuable has other properties providing such reasons⁸⁴. He promotes a view on values that he calls 'the value of choice account', where the important thing is not to choose

⁸⁴ Scanlon 2000, p. 99.

but to have the opportunity to do so, and he claims that this explains our notion of responsibility.

On the basis of this conception of value he can, instead of considering the challenge to our idea of freedom and voluntarism as being that of determinism, make the weaker claim which he calls the Causal Thesis⁸⁵: ‘This is the thesis that all of our actions have antecedent causes to which they are linked by causal laws of the kind that govern other events in the universe, whether these laws are deterministic or merely probabilistic’. I suppose this is the same as saying that we cannot accurately predict the future, which we could have done with enough information in the deterministic case. However, we can give causal explanations of what happened, and probabilistic assumptions about what is to come. Predictions about the future, I believe, are something one can achieve in controlled and idealized experiments or when dealing with ideal theories – theories telling us how things should be in ideal cases. Scanlon’s contractualism offers an account of responsibility and freedom not based on the assumption of the dualities of theoretical and practical reasoning, or for that sake of reasoning and emotions, and yet manages to deal with a world view of science more modern than the one Korsgaard’s view can meet.

I find one of her main problems to be due to her division of perspectives, upon on the dualities her view is based. To require that from a third person perspective one should aim to deal with determinism is, in my opinion, a too strong a demand. Hence the problem that Kant posed concerning the tension between freedom and laws should not have had as one of its premises that our conception of freedom needs to deal with determinism; but rather seek a scientific world view more in accordance with what Scanlon proposes, a scientific world view also dealing with the more probabilistic and spontaneous aspects of nature.

4.4 Conclusion

As I have argued, Korsgaard’s distinction between the categorical imperative and the Moral Law has lead to some problems. Her arguments confuses whether autonomous agency is equal to rational agency, moral agency or to both. Further, her constructivist approach in establishing the categorical imperative gives rise to concerns of self-

⁸⁵ Scanlon 2000, p. 250.

reference. It is hard to see how her radical constructivism can succeed in establishing a rational basis for moral obligations.

Korsgaard poses the same problem as Kant concerning how to combine the issues of freedom and law. In my opinion, one problem lies in the premises for the posed problem; the idea that nature is deterministic. Similarly, she adopts the strict command from the Formula of Humanity; that is, the command telling us that we must never treat anyone merely as means. Basically, I disagree in what she considers to be the challenges that have to be met. It is as if she is still fighting the ghosts of Kant, which are the ghosts of determinism and egoism. But why care about determinism and absolute egoism if one, unlike Kant, wishes to defend a view dealing with daily life concerns – or with the ‘real world’? Her ambitious aim of a double-level theory, capable of dealing with both practical and more ideal situations I find problematic. Her view combining autonomous agency with the belief that morality is grounded in human nature, leaves us with a dilemma. If one understands her conceptions of autonomy and egoism in a Kantian manner, then morality is grounded in rational nature and one misses some of the practical aspects which Korsgaard actually aims at. If one instead sees morality as grounded in human nature, then the question is how to justify the necessity she associates with the categorical imperative and with moral obligations. In my opinion, her double-level strategy results in taking a position that is incoherent. But why should the fact that we are the source of the law that binds us, necessarily lead to freedom and responsibility? For Kant and Korsgaard this seems to be the solution to the problems arising out of seeing nature as deterministic and having an absolutistic conception of egoism. If one did let go of these – in my opinion unnecessary – worries, autonomy is perhaps not the answer to our quest for responsibility.

Bibliography

- Allison, H. E., 1983, *Kant's Transcendental Idealism*, Yale University Press, New Haven and London
- Allison, H. E., 1996, *Idealism and Freedom*, Cambridge University Press, Cambridge
- Allison, H. E., 1990, *Kant's theory of freedom*, Cambridge University Press, Cambridge
- Bagnoli, C. 2006, "Respect and Membership in the Moral Community", *Ethical Theory and Moral Practice*, 10/2, 113-128.
- Bardon, A. 2006, "Transcendental Arguments", *The Internet Encyclopedia of Philosophy*, www.iep.utm.edu/t/trans-ar.htm, pp. 1-2,
- Cohen, G.A. 1996, "Reason, humanity, and the moral law", in Korsgaard 1996a
- Davidson, D. 2006, "Mental Events", *The Essential Davidson*, Oxford University Press, Oxford, 105-121. [Originally published 1970]
- Freeman, S. 2007, *Rawls*, Routledge, London and New York.
- Freeman, S. 2002, *The Cambridge Companion to Rawls*, Cambridge University Press, Cambridge
- Føllesdahl, D. 2005, "The emergence of justification in ethics", *European Review*, Vol.13, No. 2, 169-182.
- Goodman, N. 1983, *Fact, Fiction and Forecast*, Harvard University Press, Cambridge, MA
- Kant, I. 1997, *Groundwork of the Metaphysics of Moral*, transl. M. Gregor, Cambridge University Press, Cambridge.
- Kant, I. 1996, *The Metaphysics of Morals*, transl. M. Gregor, Cambridge University Press, Cambridge.
- Kant, I. 1997, *Critique of Practical Reason*, transl. M. Gregor, Cambridge University Press, Cambridge.
- Kant, I. 1997, *Critique of Pure Reason*, transl. P. Guyer and A. W. Wood, Cambridge University Press, Cambridge.
- Korsgaard, C. 2003, "Necessity and the synthetic a priori: A Response to Derek Parfit" Unpublished; written for a conference on Derek Parfit at Rutgers in 2003, <http://www.people.fas.harvard.edu/~korsgaard/Korsgaard.on.Parfit.pdf>
- Korsgaard, C. 1996a, *The Sources of Normativity*, Cambridge University Press, Cambridge.

- Korsgaard, C. 1996b, *Creating the Kingdom of Ends*, Cambridge University Press, Cambridge.
- Korsgaard, C. 1997a, *Introduction*, Kant 1997.
- Korsgaard, C., 1997b, 'The Normativity of Instrumental Reason', in *Ethics and Practical Reason*, G. Cullity and B. Gaut (eds.), Oxford: Clarendon Press.
- Korsgaard K. 2003, "Realism and Constructivism in Twentieth-Century Moral Philosophy." 2003 Philosophy Documentation Centre, 92-129. Available online at <http://www.pdcnet.org/pdf/8Korsgaard.pdf>
- Kvanvig, J. 2007, "Coherence Theories of Epistemic Justification", Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/justep-coherence/>
- Lewis, M. 2004, *Handbook of Emotions*, The Guilford Press, New York
- Lübcke, P. 1993, *Filosofi Leksikon*, Forsythia, p. 434.
- Nagel, T. 1986, *The view from Nowhere*, New York: Oxford University Press.
- Nelkin, D.K. 2000, "Two Standpoints and the Belief in Freedom", *The Journal of Philosophy*, Vol. 97, No. 10, 564-576
- Nussbaum, M. 2001, *Upheavals of Thought*, Cambridge University Press, New York.
- Parfit, D. 2002, *What We Could Rationally Will*, The Tanner Lectures on Human Values, University of California at Berkeley.
- Quine, W. O. 1953, "Two Dogmas of Empiricism" in from a logical point of view, Harvard University Press, Cambridge, Massachusetts
- Rawls, J. 1999, *A Theory of Justice*, Harvard University Press [originally published 1971]
- Reath, A., Herman B., Korsgaard, C. 1997, *Reclaiming the History of Ethics*, Cambridge University Press, Cambridge.
- Scanlon, T.M. 2000, *What We Owe to Each Other*, Cambridge Mass. Belknap Press of Harvard University Press.
- Scanlon, T.M. 2002a, *Rawls on Justification*, in Freeman 2002.
- Scanlon, T.M. 2002b, "Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh", *Ethics* 112, 507-528.
- Sherman, N. 1997, *Making a Necessity of Virtue*, Cambridge University Press, Cambridge.

- Skidmore, J. 2002, "Scepticism about Practical Reason: Transcendental Arguments and their Limits", *Philosophical Studies* 109, 121-141.
- Stern, R. 2007, "The value of humanity; reflections on Korsgaard's transcendental argument", Unpublished, presented at The Nature of Transcendental Arguments Workshop, <http://www.essex.ac.uk/philosophy/tpn/transarg.shtml>.
- Strawson, P.F. 1966, *The Bounds of Sense*, Methuen & Co. Ltd, London.
- Stroud, B. 1968, "Transcendental Arguments", *The Journal of Philosophy*, Volume LXV, NO. 9, 241-256.
- Williams, B. 1996, "History, morality, and the test of reflection", Korsgaard 1996a).
- Williams, B. 1973, *Problems of the Self*, Cambridge University Press, Cambridge.
- Van Willigenburg, T. 2002, "Shareability and Actual Sharing: Korsgaard's Position on the Publicity of reasons", *Philosophical Investigations* 25:2, 172-189.
- Wittgenstein, L. 1953, *Philosophical Investigations*, Blackwell Publishing, Australia.