Practical Identity and Moral Normativity

A critical examination of Christine Korsgaard's 'Sources of Normativity'

Marius Mangseth

Master's Thesis
Department of Philosophy, Classics, History of Art and Ideas

UNIVERSITY OF OSLO

May 9, 2012

Practical Identity and Moral Normativity

A critical examination of Christine Korsgaard's 'Sources of Normativity'

2012

Practical Identity and Moral Normativity: A critical examination of the Sources of Normativity

Marius Mangseth

http://www.duo.uio.no/

Print: Reprosentralen, University of Oslo

Abstract

This text explores the views presented by Christine Korsgaard in her book *The Sources of Normativity* from various angles. Her main argument is inspired by Kant and tries to establish the 'rational necessity' that at an agent must be moral. An aim of this thesis is to explicate this idea of rational necessity as well as the relation between Korsgaard's notions of 'practical identity' and 'moral identity', key concepts in the argument for morality as a necessity that reveals itself to the reflective agent.

After a summary of the book, the core argument is analysed, along with the key concept of rational necessity, the impact of the argument on reasons and justification, an illustration of the deliberative process, and finally a psychological reading of the concept of practical identity. As various criticisms are then aired, a main contention of this thesis will be that several critics read the concept of rational necessity as descriptive rather than normative for the agent. Korsgaard position is defended against such views, but criticised for failing to show how the sceptical agent is to be motivated by Korsgaard's ideal considerations.

Preface

In *The Sources of Normativity*, Korsgaard presents a Kant-inspired view of ethics, by which she maintains that the source of obligation is autonomy, meaning self-legislation by a free will. Before presenting her own argument in full, she discusses the strengths and shortcomings of different views on the source of normativity during the modern period. This provides the framework for Korsgaard's own view, conceived as a synthesis of previous views.

Korsgaard claims that rationality is not optional for a human being, and that within this rational sphere, morality is not optional, either. She claims that moral reasons for action and moral obligations arise from the powers of reflection found in every human being. While there are no moral facts 'out there', in the sense that they are fully independent of agents, there are nevertheless moral facts – or rather: rationally justified moral reasons for action – that emerge in virtue of what we are; reflective beings who confer value. The justification of morality thus issues from the concept of agency itself; the moral 'should' is inherent to what an agent *is*. The actual agent, however, must come to see this before becoming moral.

The novelty of Korsgaard's approach primarily lies with the addition of the concept of 'practical identity' to Kant's philosophy, as a way of deepening the appeal of his Humanity and Kingdom of Ends formulations of the Categorical Imperative as well as bridging them with the Universal Law formulation. This 'practical identity' pertains to the agent's sense of self, and is defined as 'a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking' (Korsgaard, 1996a, p. 101). Our deliberations and actions are governed by many such practical identities, descriptions like 'mother', 'student' or 'utilitarian'; all are carriers of normative standards that provide us with reasons. It is in light of these standards that our actions and lives are made sense of, and without *any* such practical identity to govern us we would not have 'any reason to live and act at all' (Korsgaard, The Sources of Normativity, 1996a, p. 121).

Her definition of a human being, then, is ‘a reflective being who needs reasons to act and live’ (Korsgaard, 1996a, p. 121). Practical identities are mainly contingent, though, as we are born and raised into them, or come to adapt them for various but insufficient reasons. Valuing oneself as, say, a Mafioso, may give rise to a lesser kind of reasons, but it will not give rise to moral reasons. Ultimately, however, there is one fundamental and necessary practical identity underlying all contingent identities, namely our practical identity as human beings, through which we value our own humanity as well as that in others. If we are rational, we *must* adopt this moral identity as the one to guide all others, and shed contingent identities that are at odds with the moral.

In my treatment of Korsgaard’s critics I find myself defending her to some extent, especially when the criticism is based on a misunderstanding of the concept of ‘rational necessity’, describing what the *rational* agent *must* do. This is the normative ideal of what the *actual* agent *should* do, not something he *has to do*. However, not all criticisms are as easily defeated. In seeking to establish the morally motivated agent as the only truly rational agent, Korsgaard is in danger of removing herself too far from actual human agency, even when the fully rational is considered as a normative ideal and not as actuality.

Regardless of its ultimate success or failure, it is clear that Korsgaard has presented one of the most interesting and comprehensive moral theories of modern philosophy. In what follows I will present, elucidate, defend and criticise her argument as it is found in *The Sources of Normativity*, drawing on various sources.

Acknowledgements

I would like to thank my supervisor Prof. Kjell Eyvind Johansen for his invaluable support and philosophical input during my work on this thesis. Thanks to Kjetil for being a good friend and philosopher, being there when I am in need, and to Ida for reminding me that most agents are actual agents. Thanks to mom and dad for providing me with unconditional love and support always. Finally, to my love, Anja, for loving me back even when my mind is wandering off into the abstract.

Table of contents

1 The Sources of Normativity

The Sources of Normativity is divided into nine chapters; the first four chapters are lectures in which Korsgaard presents her views, the next four are responses and criticisms by different philosophers, and finally, in chapter nine, there is a reply by Korsgaard to her critics. In my opening chapter, I will try and present the main arguments and assumptions of her four lectures in *The Sources of Normativity* as faithfully as I can, without entering into any thorough discussion of their merits. Unless otherwise stated, the claims presented in chapter one are hers, as well as I can convey them. This means omitting the criticisms of the latter half of the book for now, as well. I will return to them in my third chapter, after analysing her main argument and some of its assumptions in my second chapter.

1.1 Lecture one: The normative question

Korsgaard starts out with a question, what she calls the 'normative question':

> what justifies the claims that morality makes on us[?]
>
> (Korsgaard, 1996a, pp. 9-10)

This is really a fundamental question in modern meta-ethics, related to the common 'why should I be moral?'. To answer the question of what makes morality normative, Korsgaard says, one needs a theory of moral concepts. Some concepts she mentions in this context are goodness, duty, obligation, virtue and justice. Any theory about such concepts should contain the following trio, according to Korsgaard: *analysis* of the concepts, an account of their *application*, and an account of their *source* (Korsgaard, 1996a, p. 10).

Korsgaard goes on to remind us that the idea of morality and moral concepts are very important to us, and have a profound 'practical and psychological' impact (Korsgaard, 1996a, p. 12). In other words, morality influences the way we feel, judge and act, something which any theory of moral concepts must be able to address. This leads us to a further distinction and demand on such a theory: the distinction between *explanatory* adequacy and *normative* adequacy. The 'explanatory' part refers to the question of how morality came to play the important part that it does in human life, and the 'normative' part to the question of whether

and how this use is justified. This latter question amounts to the same as the initially mentioned 'normative question'. Before one can hope to answer the normative question, though, explanatory adequacy is a requirement. The criteria named above for a theory of moral concepts will amount to the same as explanatory adequacy. This, in turn, is necessary but not sufficient for normative adequacy.

Korsgaard illustrates the difference between explanatory and normative adequacy through discussion of a generic evolutionary theory of morality; such a theory would perhaps focus on genetic structure and see 'good' behaviour as behaviour that promotes the survival of the species. As such, the theory has explanatory adequacy, it tells us *why* we are moral creatures. However, this does not have normative adequacy: when we are deliberating about what to do, the theory does not provide us with reasons why we should act morally. Biology does not justify moral action.

Let's say that I am deliberating about whether to turn in the Jews I am housing to the Nazis on my doorstep. Should I risk my life to save them, because that is what morality demands of me (which we now assume). Being an believer in the standard evolutionary theory of morality, I know that it is 'only biology' that presents me with an urge to save them for the preservation of the species – but is this really a good reason for me? On second thought, I might be better off fighting this urge. After all, devoid of moral reasons that seem authoritative to me, I might look to what is otherwise in my best interest. And risking my life for strangers might not be in my best interest at all, seen from an egoistic point of view. Thus, I might not find it best or rational to act morally even though my natural inclination is to be moral, and the case may very well be that my moral impulses are worth fighting.

Korsgaard's view is that a theory about morality that does not support morality is normatively inadequate. It may explain why we tend to act morally, but it does not have the normative power to convince anyone about to make a moral choice, doubting whether they are really required to do the right thing. Korsgaard points out that while explanatory adequacy is a third-person affair, normative adequacy is a first-person issue (Korsgaard, 1996a, p. 16); the theory must provide the reflective individual with some justified reason to be moral in order to answer the normative question. This is why the difference is crucial between ethical theories who seek explanatory adequacy alone and those who try to come up with a justification of

morality as well. Korsgaard is a philosopher of the second type – she is out to show that morality's claims on us are warranted, rational and not a matter of taste.

For a theory to have normative – or, as Korsgaard sometimes calls it, *justificatory* adequacy, three criteria are introduced:

> 1) The theory must provide an answer that satisfies the person who asks himself the normative question in the first person – the answer must be an answer for *me*.
>
> 2) The theory must have 'transparency' – that is, it must be such that if we were to know the real reasons for our being morally motivated (according to the theory), us knowing these reasons would give us reasons to sustain rather than question and possibly give up our moral practices.
>
> 3) The theory must 'appeal, in a deep way, to our sense of who we are, to our sense of identity'. This entails that the theory must show that 'sometimes doing the wrong thing is as bad or worse than death'. Thus, the theory must appeal to our sense of integrity; that in doing the wrong thing we would not be ourselves anymore, we would face loss of identity, something which according to Korsgaard can be just as bad or worse than dying.
>
> (Korsgaard, 1996a, pp. 16-7)

At this point, Korsgaard's basic criteria of a successful theory of morality have been laid out, and by introducing the third criterion, she includes the concept of personal identity as central to her approach. This indicates that a theory of what one should do must be underpinned by a theory of how we view ourselves, who we are, and not in any sense; a criterion of success for a moral theory is that it should be able to inspire you to lay down your life, or at least to consider it, for the sake of moral obligation.

Having laid down these strong demands on successful moral theories, Korsgaard goes on to address different ways that the moral question has been answered throughout modernity. The four approaches she presents are seen as historically successive, one view developing as a response to the shortcomings of the previous one. The approaches to grounding moral normativity that are addressed are voluntarism, realism, reflective endorsement and autonomy. According to Kant and his followers, *autonomy*, i.e. the self-legislative capacity of a free will, is the source of normativity. We have the power of rational reflection, which

enables us to lay down rules, or laws, to govern our own behaviour as well as that of others. Korsgaard's aim is to show that some version of Kant's view is the ultimate account, or at any rate the best available account of the source of normativity. I am not yet sure how strong the claim is. According to Korsgaard the Kantian view entails, yet surpasses the previous views and their shortcomings, so that she ends up saying that there is some truth in all of the accounts, but only in the light of Kant. Before reaching this conclusion, however, there is a lot of criticism to be made of the previous views, in order to create a kind of historical narrative leading up to Kant, or rather, Korsgaard's version of Kantianism. Now, let us explore the relation between these views together with Korsgaard.

According to *voluntarism*, the source of normativity is some authoritative will, be it the command of God or a Hobbesian sovereign[1]. In modern times, i.e. from Hobbes and onward, the contractarian view is the prevalent one, as God and the idea of a teleological account of the world have given way to a scientific world view. The modern voluntarists, like Pufendorf and Hobbes, accepted that emerging scientific world view. This, in turn, lead them to believe that values can't be found in nature, or only 'found in nature' to the extent that the will of persons put them there. Insofar as this will is human and not divine, this is an anti-realist view of morality. The label is mine, Korsgaard doesn't explicitly write about anti-realism. This may be on account of her self-titulation as a moral realist, that is, a *procedural*, and not *substantive*, moral realist (Korsgaard, 1996a, p. 112). For her this means that in some sense, obligation is real and binding and that the procedure of giving ourselves laws produces it. As far as I can tell, the realism that the anti-realist opposes is the substantive one, and as such, Korsgaard might be some sort of anti-realist herself, but at any rate calls herself a procedural realist in order not to be lumped together with those anti-realists with less confidence in the justification of morality, like *noncognitivists*[2] and *moral error*[3] theorists. What this shows is mainly that anti-realism can be a confusing term, and this may be why Korsgaard avoids using it in the first place.

At any rate, Hobbes' approach, while taking the emerging scientific world view into account, runs into some trouble; we are obligated by the laws of the sovereign to whom we have

[1] A regent whose will is the law on account of some social contract, actual or theoretical.
[2] The view that moral judgments have no truth-value.

[3] The view that they do have a truth-value, but that they are systematically false.

conferred power, as part of a practical arrangement where we give up some of our personal freedom in order to be able to live in less fear of one another. But our obligations spring from the power and wisdom of the sovereign, none of which are absolute. It seems we are only obligated to the extent that he is able to catch and punish us and/or to the extent that his laws are perfectly wise and benevolent; the first implies that we aren't wrong as long as we don't get caught, and the latter appeals to some standard of wisdom and benevolence that would require further justification. Basically, the obligation to the lawgiver seems contingent, and it seems we can still question why we should be obligated by him. In Korsgaard's words: 'the very notion of a legitimate authority is already a normative one and cannot be used to answer the normative question' (Korsgaard, 1996a, p. 29).

According to *realism*, there are intrinsically normative entities or facts which give truth-value to normative propositions. In the realist camp, Korsgaard places Clarke, Moore, and more recently, Nagel. The realist response in order to stop the voluntarist's looming regress is to postulate the existence of objective values, reasons, obligations, or actions, i.e. entities that are intrinsically normative. However, the realist's belief that they are irreducible and that it is a mistake to try and explain them might not be very convincing. Although we might at first glance agree that some actions seem to have to-be-doneness built into them, or that some situations scream out for us to take action, like a rape in progress or a baby crying, this doesn't really satisfy the reflective agent who recognises the claim of morality but doubts whether it is justified. It seems that the realist doesn't really stop the regress; he just wants us to stop all questioning by declaring that moral facts are simply there. Rather than answering the normative question, this approach just shuns it, according to Korsgaard. During the discussion of realism, she returns to the issue of explanatory and normative adequacy. Sometimes, the one is mistaken for the other, and this leads to confusing results. She employs Moore's 'open-question argument' to illustrate this (Korsgaard, 1996a, p. 43). Moore argued that however we analyse the concept 'good', it is always relevant to ask whether the objects picked out by the analysis are actually good. Thus, Moore thought, 'good' is an unanalysable concept, and we should simply intuit what is good. However, Moore's argument actually rests on the importance of the normative question; all that he really points out is that when we have some *explanation* of the concept, *justification* is still required to satisfy us. So we should not believe that this renders normative concepts indefinable, only that there are further questions to settle before a proper moral theory is in place. Korsgaard argues that such confusion is

typical of realists. Furthermore, John Mackie's well-known 'argument from queerness' discredits realism by arguing that the intrinsically normative entities must be strange entities indeed, of a different kind than anything familiar to us (Mackie, 1977, p. 38ff). Korsgaard counters Mackie in an original way at the end of her final lecture by stating that 'it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. They are people, and the other animals' (Korsgaard, 1996a, p. 166). Like I mentioned before, this is Korsgaard brushing off the *substantive* moral realist, whereas she considers herself to be a *procedural* moral realist. For the substantive moral realist, the moral truths are simply 'out there', while for the procedural realist, they are the result of a correct procedure for arriving at moral truths. This does seem to be the same as saying that there are moral reasons and moral truths, and that they are in some sense mind-dependent. During her fourth lecture, Korsgaard apparently denies the mind-world dichotomy; her claim, as far as I can tell, is that moral goodness and badness exist in the world, but take a perceiving rational animal to identify (Korsgaard, 1996a, p. 155). Identifying moral value, then, requires the human, or rational, perspective. This is probably consistent with Kant's transcendental idealism.

1.2 Lecture two: Reflective endorsement

According to proponents of the *reflective endorsement* strategy, like Hume, Mill and Bernard Williams, the source of morality is in human nature, and once this has been properly explained, we can endorse or reject the claims of our nature depending on whether we still believe morality is good for us. Nietzsche, for one, did not believe it was – at least not the kind of morality[4] that philosophers in general tend to try and justify. His thought leads one to question one's moral inclinations, just like the evolutionary theory in the example I presented earlier, though Nietzsche seems to leave more room for nurture to play a role. Whether a consequence of our nature or nurture, morality might be no more than a source of bad conscience that we had better shrug off for the sake of our own well-being.

[4] Korsgaard's use of the term 'morality' is not very precise; sometimes, It is reasonably understood as the contents of a general but unspecified agreement on what's right and wrong; sometimes, her use refers to her own normative conception of morality, implying that obligation is real, that we sometimes have moral reasons to do what we don't want to do. At other times, 'morality' refers to the idea of morality implicit in the theory of moral concepts in question. However, this unclarity is no major source of confusion.

To return to those first mentioned, who are slightly more optimistic about preserving morality; like Nietzsche, they have understood the challenge of the normative question – perhaps *the* modern problem – and correspondingly, they have tried to come up with a theory that can survive critical reflection. Reflective endorsement basically means aiming at normative adequacy; first, one worries that the true account of our moral motives and beliefs may not be one that sustains them, and then the task is clear: to come up with an account that is endorsable even when the normative question is pressed to the full. Hume and Williams, according to Korsgaard, reject the substantive realist's claim that objective moral values are out there, viewing moral properties as projections of human sentiments onto the world (Korsgaard, 1996a, p. 50). Mill, to whom we will return, may see pleasure/pain or desire as intrinsically normative, but this is not essential in this context, as he makes a stab at further justification anyway. Moral philosophy, then, should have facts of human nature as a starting-point, and not concern itself with intrinsically normative entities somehow prior to the human outlook.

At this point, Korsgaard is not out to criticise the method of reflective endorsement like she did voluntarism and substantive realism. After all, the method is related to the standard of normative adequacy introduced at the beginning of her lectures. The problems that ensue when accepting an evolutionary theory of morality, for instance, are related to its failing a test of reflective endorsement. Korsgaard's conclusion in the second lecture is that autonomy is the only explanatory concept that can warrant full endorsement. The principal difference between Kant and the others seems to be that he can account for reflective endorsement of each particular action, where Hume's and Williams' accounts stop at the general level of endorsing dispositions, running into trouble in some concrete dilemmas. For now, let's explore Hume's moral philosophy through Korsgaard's eyes.

Hume's moral theory is a theory of human sentiments. Critical of the realist approach, he observed that the inherent wrongness of a wrongful action was nowhere to be seen in the circumstances, but rather that the wrongfulness was inherited from our condemnation of the action in question, and this, in turn, is a question of our natural disposition to feel revulsion towards heinous acts and sympathy and warmth towards acts of benevolence and the like. The normative question becomes whether these natural inclinations of ours should be endorsed. What standard, what perspective should we apply in order to decide whether we should

endorse or reject this nature of ours? This is what Korsgaard means when she talks about the choice of a point of view. There are several normative points of view that can seem appropriate for the task at hand. For Hume, the important ones are self-interest and the point of view of morality itself (Korsgaard, 1996a, p. 55). Hume argues that our natural tendency when making moral judgments is to applaud qualities that are *agreeable* and *useful* to ourselves and to others. General qualities or virtues that contribute positively to our usefulness and agreeableness result in a feeling of pride on our part, thus contributing to our own happiness. Conversely, lack of such qualities results in humility, according to Hume a lack of self-worth that is detrimental to our happiness. Now, these moral sentiments that cause us to feel well when we are useful and agreeable to *ourselves*, are self-evidently to be endorsed from the point of view of self-interest. When it comes to being *agreeable* to *others*, Hume similarly argues that being liked and being benevolent are better for our pride – and consequently, our happiness – than being despised and selfish. As for being *useful* to others, the argument is as follows: any kind of desire-fulfilment contributes in some way to our happiness by way of satisfaction, but benevolent desires have an immediately 'smooth, tender and agreeable' feel to them (cited in Korsgaard, 1996a, p. 57). Perhaps what Hume is saying is that they are less controversial, on account of their being conducive to our own agreeableness, both in the eyes of self and other. Thus, he concludes, we should endorse our moral sentiments in full from the point of view of self-interest. Of course, there might still be the question of a situation where others are not present to catch or condemn us and where the action doesn't seem to undermine the system of justice, which it is in our interest to uphold. Why should we be virtuous *then*? This is what Hume's 'sensible knave' asks himself. The answer is already implied in what Hume has said about being agreeable in our own eyes, as well as the mere thought about what others *would* think if they saw us. So it turns out that even in the case where we are lacking an immediate selfish motive to be virtuous, we have upon reflection a long-term interest in our own integrity, to be consistent in our own eyes and not experience the mixed feelings of instant gratification and humility. What Hume would probably tell the knave is that although his actions won't hurt the system of justice, such behaviour will nevertheless give him a sense of humility that's detrimental to his happiness in the long run. We have a duty to ourselves to be moral, grounded in Hume's account of our human nature viewed from a self-interested point of view. Hume calls this obligation from the point of view of self-interest 'interested obligation'. I suspect that this entails what Harry

Frankfurt calls a second-order volition[5], in this case a duty to have one; Hume says that upon reflection, it is in our interest to be agents who want to be virtuous. Even when we have no desire to be virtuous, we should, if we believe Hume's theory, want to have this desire for the sake of our happiness. If the knave were virtuous in the first place, his problem wouldn't arise, because then he would desire to be virtuous, and acting virtuously would be its own reward. And cultivating this desire to be virtuous is a matter of repetition, of habit, of upbringing.

The harmony between the point of view of self-interest and the point of view of morality, then, is established. Hume's account basically says that upon reflection, egoism and morality are not in conflict, i.e. they are what John Rawls calls 'congruent' (Korsgaard, 1996a, p. 60fn27). Even so, there is something amiss. Even though morality is not detrimental to our happiness, many will say that a moral theory supported by an egoistic perspective is no proper moral theory. According to realists, we should be moral for the sake of being moral. This brings us to whether our moral sentiments can be endorsed from the point of view of morality itself, in other words: whether we can find an intrisic justification of morality as well as the extrinsic we just established. Korsgaard believes that Hume manages this through what he calls the 'reflexivity test', a test of the normativity of a given faculty. 'When the faculty takes itself and its own operations for its object' and 'gives a positive verdict' (Korsgaard, 1996a, p. 62), one can trust its verdicts in general. This, according to Hume, is the case with the moral sense: it approves of itself upon reflection. So what is it about the moral sense that makes us feel more confident about it when we reflect on it and 'those principles, from whence it is deriv'd' (Hume cited in Korsgaard 1996a, p. 63)? Apart from that which has already been said about self-interest, it is somewhat unclear to me whether there are further reasons why the moral sense approves of itself. Neither the Hume citation in full nor Korsgaard seems to list them. Nevertheless, I will make a stab at getting Korsgaard's point across. Korsgaard makes a number of related claims: there are only a limited number of normative points of view to which we can appeal; outside normative points of view, normative questions can't be asked in a meaningful way; outside human nature, there is no point of view from which

[5] A desire for our will to be determined by some desires rather than others. The ability to have second-order desires is according to Frankfurt characteristic of the human will (Frankfurt, 1997, p. 16). Korsgaard's general account of the will seems to harmonize with Frankfurt's, although he does not argue for any special authority of moral considerations.

morality can be challenged; internal to human nature, morality can be challenged by self-interest (among other things) and vice versa (Korsgaard, 1996a, pp. 64-6). So the question here is, can morality be meaningfully challenged by itself? The answer is no, but as long as it can be unsuccessfully challenged by other points of view internal to human nature, like self-interest, that is as far as we can come. This does not seem to answer Prichard's realist challenge that 'if a question admits only answers that are circular [i.e. a moral justification of morality] or irrelevant [i.e. a egoistical justification of morality], then it is a mistake to ask that question' (Korsgaard, 1996a, p. 32). But perhaps what Korsgaard is getting at is what we have already explored, that Hume establishes a kind of duty to be moral – 'interested obligation'. Hume argues for the intrinsical normativity of human nature, at least. The motive for being moral should be morality itself, according to Prichard; that we ought to be virtuous in general because it is a part of our best nature to be moral is perhaps just this. But this 'ought' and 'best' still refers to a positive evaluation of our moral sense from the point of view of self-interest. Korsgaard finally draws on Shaftesbury, speaking from a voluntarist tradition when he says that our nature has authority over us by virtue of its ability to punish us. Korsgaard claims this is, like Hume's argument, different from mere reference to self-interest, since it is then a question of a motive of duty (Korsgaard, 1996a, p. 66fn). Hume's point, according to Korsgaard, is that it is in our interest to be people who practise virtue for its own sake (Korsgaard, 1996a, p. 60). And this implies that we should be moral even when we don't want to be moral, which is the nature of duty. The distinction then, might be that Hume's view indicates that it is in our self-interest to be moral from an objective third-personal point of view, which can be endorsed by an agent from his first-person point of view upon reflection, even when not in his immediate interest before reflection, like in the case of the sensible knave. The moral sense has authority in that it provides us with a motive of duty, pointing us towards the interests we should have for our own good, not the ones we happen to have. Consequently, Hume would scorn what Korsgaard calls the wanton, the 'slave of the passions', but his view is still compatible with the egoist, the 'steward of her own interests' (Korsgaard, 1996a, p. 101). So all in all, Hume might not satisfy Prichard, but at least he provides us with a way saying that we should practise virtue for its own sake. Unless Korsgaard's own way is acceptable, perhaps something like Hume's approach is the way to approach the normative question without appealing to intuition or the like.

Korsgaard goes on to treat Bernard Williams and John Stuart Mill in terms of their use of the reflective endorsement method. Williams is likened to Hume in that he finds that morality must be established through congruence with human interests, but in his case, it is a question of human flourishing rather than self-interest. Williams takes note of the fact that there are many different moral practises both across cultures and inside them. In order to evaluate any given moral practise, the normative question becomes: is the social world where those values prevail a good place for human beings to live? As for what characterises human flourishing, Williams awaits a comprehensive study, including insights both from social and natural sciences. This leads us to a difference between Hume and Williams that Korsgaard emphasises: in the nature vs. nurture debate, Williams is more on the nurture-side of the argument. For Hume, although nurture plays some role, our basic moral sentiments spring from human nature. If they are not congruent with self-interest, we have a problem, since it would not be in our best interest to do the moral actions that nature has created us to feel like doing. A less human nature-dependent view of moral sentiments like Williams' sees those feelings and consequently our moral judgments as more malleable. This means that the threat diminishes in Williams' case: science coming up with a picture of human flourishing that is at odds with the values of most or all cultures, only implies the problem of changing those values – and thus the cultures – to better provide for human flourishing. Williams' view implies that when cultures are confronted with one another, the confrontation might make the members apply the method of reflective endorsement, consider the value of their values up against what they come to see as the best life and consequently discard or annex moral values or practices[6].

During the treatment of Hume, I ran into some trouble accounting for Korsgaard's emphasis on the moral sense's approval of itself. Korsgaard finds a similar element of reflexivity in Williams' thought; Williams writes that we cannot escape thinking about morality from a moral point of view, whichever one that is already a part of us (cited in Korsgaard, 1996a, p. 77). Naturally, this morality is an independent source of reasons, however relative and unjustified. It is also incapable of criticising itself from its own point of view. On Williams'

[6] I believe that this view also has a further implication: if one is optimistic – perhaps naïve – about human reflective capabilities and the best values being the ones to survive rather than the values of the powerful, values would upon Williams' view improve over time. They would also converge, given that there is just one set of 'best values'. This seems to be in the vein of Hegel, where spirit is constantly transcending itself (in the sense of 'going beyond'), becoming more complete.

account, unless the agent in question is already a brought up and convinced, full-fledged egoist, he cannot but argue from a point of view that is in some way independent of egoism, i.e. he potentially has a view that we might call moral only for the sake of being moral. However, this in no way justifies this contingent morality, which must be reflectively tested. In Williams' case, this means testing whether it is congruent with human flourishing.

So, do both Hume and Williams have some egoistic view of morality, in that they seek to find justification of morality on account of some tale about what is really in our best interest? The Prichard criticism that justification of morality from self-interest might not be the right kind of justification still seems to have some relevance. But then again, the view of Williams might be as far as philosophy can take us, as suggested by the title of his seminal work *Ethics and the Limits of Philosophy* (Williams, 1985).

Unlike Hume and Williams, Mill is a kind of substantive realist, according to Korsgaard, in that he thinks that the pleasant is valuable in itself, but he is nevertheless a representative for the method of reflective endorsement, since he also thinks that further justification is needed in order for people to see that the claims of utilitarianism provide them with moral reasons. Korsgaard also pegs Mill as an *externalist* with regard to *moral motivation*, as opposed to herself and most of the others she discusses. If correct, this means that contrary to *internalism*, he doesn't believe that a moral judgment necessarily provides motivation. If he did, he would think that his proof of the principle of utility[7] could do the job on its own. But according to Mill, what motivates us to do our duty are the internal sanctions that accompany moral beliefs, i.e. a conscience instilled in us from childhood, rather than the moral beliefs directly. This means that if we acquire new beliefs about what it is right to do, like Mill's principle of utility, then they might not necessarily motivate us at all even if we are convinced by them. Korsgaard notes that her normative question, understood as 'should we allow ourselves to be moved by the motives which morality provides?' (Korsgaard, 1996a, p. 81), takes the truth of internalism for granted, but can be rephrased in externalist terms: '…moved by such motives as may be provided for morality (either by nature or by training)?' (Korsgaard, 1996a, p. 82). The maturing individual will come to ask himself this question, according to Mill, and might find that the moral motives instilled in him are wholly arbitrary, weakening the associated

[7] The principle of utility: Actions are right to the extent that they promote happiness, understood as pleasure and the absence of pain.

internal sanctions over time – luckily, though, a utilitarian conviction will only strengthen itself upon reflection. Mill cites the 'social feelings of mankind' as a 'basis of powerful natural sentiment' that harmonises with a utilitarian view of ethics (cited in Korsgaard, 1996a, p. 83). Thus, our natural social sentiments and our reflective powers will work against any other view than the utilitarian over time. Although such congenial feelings might not be as strong in most of us as that of selfishness, we will upon reflection come to see them as feelings that it is not well for us to be without, and as feelings that sanction utilitarianism.

This account of the reflective process that takes place in the maturing individual seems to indicate that arguments can indeed motivate after all; if not directly, then over time in a slow-changing, but nevertheless malleable individual that is not condemned to a certain set of moral sentiments and motivations by their upbringing – much like Williams' moral agent. Korsgaard thinks that there is some sort of contradiction in this if Mill is an externalist (Korsgaard, 1996a, p. 85). An externalist believes that an argument doesn't necessarily motivate by itself. This is perhaps why he refers to the natural social sentiments; but if these are to strengthen themselves over time upon reflection about the truth of utilitarianism, it seems like reflection has some power to motivate after all. Or alternatively, that this reflection must be understood as a reason for those who are relevantly disposed, a reason to implement utilitarian policies that instil utilitarian moral sentiments. Perhaps what Mill believed is that even convincing arguments do not automatically make their conclusions the principles of action, which seems quite reasonable to me, if there are strong and contrary feelings long cultivated in us. So Mill might have reasoned that people would be convinced by his argument, but not act according to it immediately; rather he may have hoped that some people would aspire to become what he has convinced them is the best way to be. Although arguments can possibly motivate on a 'soft' externalist view, training and education is required as well, not one or the other like in Korsgaard's rather strongly formulated dilemma (Korsgaard, 1996a, p. 85). We might imagine politicians, convinced by Mill's argument, applying utilitarian policies without themselves being motivated utilitarians. Korsgaard thinks that Mill's argument doesn't explain why anybody should become utilititarians, it only shows that 'if there were any utilitarians, then their morality would be normative for them' (Korsgaard, 1996a, p. 85). As far as I can tell, this is taking a more severe version of internalism for granted, that one cannot coherently fail to do one's duty once the duty is recognised. I don't really think this is Korsgaard's position, so to make sense of her criticism

of Mill, I see it as emphasising his lack of faith in the power of argument, rather than an all-out moral judgment vs. upbringing dilemma. The proper place for justification, Korsgaard maintains, is in the argument, although the right upbringing surely makes things easier for the agent on her account, too.

Korsgaard concludes her treatment of Mill by saying that the normative question must address the agent that asks it, and that Mill fails to do this. If Mill's argument is convincing, though, I believe he does this to some extent, even if it only addresses the agent in the sense that he might want to become a utilitarian, i.e. not necessarily and immediately make him *be* a utilitarian that acts from utilitarian principles in whatever moral dilemma he finds himself. I don't take the 'immediate conversion' thought to be Korsgaard's description of actual human agents either, but she has high hopes for the ideal, rational agent at least being capable of such immediate motivation. As we can see, the issue of internalism and externalism about moral motivation is a delicate one. However, pursuing it further here will lead us away from Korsgaard's text, so let's leave it there for now.

Korsgaard closes her second lecture by critisising Hume and Williams for using the reflective endorsement test as a philosophical exercise, a test proving the normativity of moral dispositions and sentiments. 'The reflective endorsement test', however, 'is not merely a way of justifying morality. *It is morality itself*' (Korsgaard, 1996a, p. 89). Their approach, it seems, is too general; endorsement of certain dispositions might not resolve the agent's doubts about what to do in a particular situation where our dispositions and our moral judgment are suddenly at odds. In such a situation we would wish that we didn't have our general dispositions, because they motivate us to do what we judge to be wrong there and then. To resolve these issues, we must turn to the third lecture and the views of Kant and Korsgaard. To avoid confusion I will refer to them as Korsgaard's views, as I am not particularly interested in whether she stays true to Kant.

1.3 Lecture three: The authority of reflection

Korsgaard starts out by making some claims:

i) 'Autonomy is the source of obligation'

ii) 'We have *moral* obligations, (…) [i.e.] obligations to humanity as such'

iii) 'if we take anything to have value, we must acknowledge that we have moral obligations'

(Korsgaard, 1996a, pp. 91-2)

These claims are to be validated in the course of the third and fourth lectures. We begin with the first claim, a general one about the source of all obligations, not just the moral kind. The source of normativity is our autonomous will. Autonomy in Korsgaard's sense means self-legislation; the autonomous agent makes his own laws, according to which he is then bound to act. Autonomy also requires that the will is free. Seen from a scientific or deterministic point of view, this is not the case. However, Korsgaard argues, when seen from a first-person perspective, it is a fact about *what it is like* to be a reflective agent that our will is free. We can't help but deliberate about what to do, and in this process the will is experienced as an unbound first cause in a chain of events. The point is not to challenge determinism. However, when looking at the structure of reflective consciousness, the truth of determinism becomes irrelevant. In this first-personal sphere of practical reason, then, the will is free. So it seems that Korsgaard endorses a kind of compatibilist view.

We fail to be free if the source of our judgment is external to the will itself, including our own desires. This means that our judgments must be the outcome of a process of reflection, through which we come up with our own reasons. That we have a reflective consciousness means that we are able to question our perceptions and – more importantly here – our desires. When faced with a desire or the like, the agent can deliberate about whether or not to act on this desire. The reflective endorsement of a desire provides us with a reason. Correspondingly, a rejection provides us with an obligation (Korsgaard, 1996a, p. 102). An obligation then, is the real test of character, because we have an impulse to do something that we think we shouldn't do. After a reflective process, the source of judgment becomes our own rational will, determining itself. The reasons come in the form of principles, i.e. laws, according to which decisions are made and action is carried out. So the free will must give itself dictates that it sticks to. Autonomy means being a law unto ourselves, and this is what the Categorical Imperative, in the Formula of Universal Law[8], tells us to do: make our own

[8] 'handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde' (Kant, 1785, 4:421) – act only according to that maxim [principle] by which you can at the same time will that it should become a universal law (trans. in Korsgaard, 1985). I will follow Korsgaard in addressing this as the 'categorical imperative' without capitalization, implying only to have to choose on account of some principle, although traditionally there are several versions, including what is here referred to as the 'moral law'.

laws. The categorical imperative is the principle of the free will, derived from the very nature of this will; having a will means having to choose whether to act on a given desire, and needs some law, some standard to do so.

Given that Korsgaard argument is correct, what has been established so far is that autonomy is the source of normativity [claim i)]. But strictly *moral* reasons, not just normative ones, are what we are after, and the categorical imperative doesn't say much about the content of the laws. To get from normativity in general [claim i)] to moral normativity [claim ii)], however, there are a few steps to be made. Korsgaard begins by making 'a distinction that Kant doesn't make' (Korsgaard, 1996a, p. 98). She separates the categorical imperative from the *moral law*. The moral law is the law of the Kingdom of Ends, Kant's utopian republic where everybody acts as rational beings would; according to his conception of rationality, rational beings are moral beings. The moral law entails acting only on 'maxims that all rational beings could agree to act on together in a workable cooperative system' (Korsgaard, 1996a, p. 99). The moral law is a specification of the *domain* of the categorical imperative, one of several possibilities. Among these possibilities we find the law of a wanton, whose law is to act on the desire of the moment, and that of an egoist, whose law is to maximise his self-interest in the long run (Korsgaard, 1996a, p. 99). The moral law, on the other hand, means all rational agents working together, valuing one another as ends-in-themselves, and the agent thinking of himself as a member of the Kingdom of Ends. The transition from the categorical imperative to the moral law requires further argument.

This is where Korsgaard introduces the central concept of practical identity. Practical identity is defined by Korsgaard as 'a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking' (Korsgaard, 1996a, p. 101). There are many possible practical identities, and there are many of them per person; roles such as man, mother, Muslim, doctor and so forth, all providing the agent with reasons, obligations and a sense of self. Korsgaard's concept of practical identity, as I understand it, denotes sources of integrity of the self, psychologically speaking. Obviously, some practical identities are more central to one's identity and integrity than others, that is to say: some roles and their obligations and reasons can easily be shrugged off without particular consequence, whereas others are regarded as crucial. When essential roles are challenged, this can lead to an identity crisis if we fail to observe the obligation that

springs from this core identity of ours. Korsgaard thus ties the concept of obligation closely to the concept of practical identity:

> It is the conceptions of ourselves that are most important to us that give rise to unconditional[9] obligations (…) An obligation always takes the form of a reaction against a threat of loss of identity
> (Korsgaard, 1996a, p. 102)

This means that if we fail to observe the prescriptions inherent in certain crucial roles, and the transgression is sufficiently serious, we no longer know who we are, and we might as well be dead.

These crucial roles, of course, differ from person to person, as we are not all the same. The categorical imperative 'choose a law!' applies in light of the roles we happen to possess. So different laws hold for different people, and the obligation that follows from these identities is not moral in the sense that Korsgaard is looking for; role pluralism implies value pluralism and leads to some kind of relativism – not to the moral law. What Korsgaard is getting at, however, is that there might be some fundamental and unsheddable role that serves as a basis for all others, and a role that we all share and that has at least some moral implications. To complete the argument, Korsgaard seeks to show that we cannot but view and value ourselves as *human beings* in order to be agents and persons at all, and that this role as a human being is identical to that of a member of the Kingdom of Ends; a moral, rational being.

What seems likely so far is Korsgaard's claim that 'you must have *some* conception of your practical identity, for without it you cannot have reasons to act' (Korsgaard, 1996a, p. 120). Practical identity then, is the yardstick of the reflective endorsement test: reflective endorsement or rejection takes place with one or several practical identities in view. And this turns out to be exactly what the role of a human being entails: being one that has to have some practical identity, without which he cannot deliberate. Deliberation is something that is inherent to the structure of reflective consciousness. We enter a deliberative process when we take a step back and employ our ability to question our impulses before we act on them. Our practical identity enters into this in that it carries with it some conception of the good, against which proposals for action can be measured. We can, on a higher level, question the value of a given practical identity. After all, we are born into a contingent bundle of such identities, as

[9] All obligations are unconditional in form (Korsgaard, 1996a, p. 103). As we will se, such obligations are not necessarily *moral*.

well as the ones we shed and adopt in the course of our lives. So which practical identities should we have, then? In turns out that on Korsgaard's account of reflective consciousness and practical identity, the non-contingent element is that we need some such conception in order to deliberate and to have reasons at all. Recall that a human being is here defined as 'a reflective being who needs reasons to act and live' (Korsgaard, 1996a, p. 121). The claim then, is something like this: if you are to have any reason to live and act, you must have some practical identity to guide you. This in itself is a reason for having at least one practical identity, a reason that is provided by your humanity. In a sentence: in order to value anything at all, you must value your identity as someone who can and does value – you must value your own humanity. Which brings us to the third claim; 'if we take anything to have value, we must acknowledge that we have moral obligations'. In other words, valuing your humanity must mean having moral obligations, according to Korsgaard. The second claim states that 'We have *moral* obligations, (…) [i.e.] obligations to humanity as such'. So, in order to get to moral obligations, what remains to be demonstrated is that valuing your own humanity is connected to valuing that of others, 'humanity as such'. At this point, Korsgaard just states her belief that valuing your own humanity rationally requires valuing that of others (Korsgaard, 1996a, p. 121), and promises to return to the matter in lecture four.

At any rate, what the argument intends to show is that one cannot avoid identifying oneself as a human being, since it is a practical identity we cannot reasonably choose or reject. That doesn't mean that all agents already and explicitly value themselves in the way conceived by Korsgaard, but it means that they are so committed by reflecting on the nature of their agency.

It is an account of way we are, from the point of view of rationality, and not from biology, like Hume's starting-point. With the aid of reflection and Korsgaard we see this to be true. Korsgaard says her argument is 'transcendental' (Korsgaard, 1996a, p. 123); as such, it is an investigation into the conditions that make rational action possible; it aims to show that rational action is only possible insofar as humans value themselves. Rational action exists[10], Korsgaard says, so we know that it is possible (SN 124). Rational action requires us to value something. Valuing something requires us to value ourselves as the conferrers of that value. So, since there is rational action, which Korsgaard takes as a fact that everybody will agree

[10] Whether it is actually instantiated is not empirically verifiable, though. Korsgaard's argument can be seen as an investigation contingent on the possibility of such instantiation. See the quote in my Conclusion.

on, human beings must find themselves valuable, since it is a precondition of rational action. This further means that human beings *are* valuable from a first-person point of view, that is, from the point of view of practical reason or reflective consciousness (Korsgaard, 1996a, pp. 123-4). In this last step from finding ourselves valuable to actually being valuable, there is an obvious similarity with the argument for freedom of the will found at the beginning of the lecture. Both our value and our freedom are facts, but not from a third-personal, scientific point of view;

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

Korsgaard believes she has found the necessary standard that grounds a universalist ethical theory like her own, the standard that unites us and stops further questioning about what to do. Valuing yourself as a human being is what the otherwise contingent practical identities have in common. Yourself as a human being becomes the governing, necessary identity that confers value onto contingent identities. As such, it serves to settle any morally relevant disputes between them, being a standard against which they can be measured. But there is still important work to be done. Most importantly, Korsgaard must validate the claim that valuing your own humanity rationally requires valuing that of others. Further, she should address the contingency that we might not take anything to have value. Those are some of the topics of lecture four.

1.4 Lecture four: The origin of value and the scope of obligation

In order to prove that I have reason to value the humanity of others, Korsgaard asks whether reasons are of a private or public character. If she can show that there is no such thing as private reasons, she will have come a long way towards the conclusion she is looking for; if I value my humanity for my reasons, and those are reasons for you to value my humanity as well, and vice versa, we are rationally required to value the humanity of one another. Korsgaard's basic claim regarding this is that 'To act on a reason is already, essentially, to act on a consideration whose normative force may be shared with others' (Korsgaard, 1996a, p. 136). To make this credible, she turns to Wittgenstein's private language argument

(Wittgenstein, 1953, pp. §244-271). On Korsgaard's reading, what Wittgenstein is saying is something like this:

Let's say meaning is an essentially private issue. If I name some sensation of mine 'S', a sensation only identifiable by me, how would I know whether I was right in my belief that I am having this sensation on future occasions? If meaning was private, there would be no criterion of correctness, because if you think you are having 'S', then you will be right no matter what. And if there is only the possibility of being correct, then you cannot talk about rightness at all. Meaning, it turns out, is essentially public and is established through a relation. Meaning is relational because it is a normative notion; to say that X means Y is to say that one ought to take X for Y. It takes two to make a meaning, a 'legislator' and a 'citizen' who obeys. If the citizen disobeys or misunderstands, he can be corrected by the legislator, the originator of the standard.

Korsgaard believes this applies analogously to the normativity of practical reasons as well; To say that reason R is a reason for action A is to say that one should do A because of R. This is also a relation between a lawgiver and a citizen who obeys, and not a causal one, in that the citizen can fail to perform A because of irrationality or defiance. Like with meaning, it takes two to make a reason. However, the two in Korsgaard's relation are not separate agents, but the two parts of reflective consciousness: the *thinking* and the *acting* self (Korsgaard, 1996a, pp. 137-8).

The point is not that private language or private reasons are impossible, but rather that any language and reasons must be conveyable to others if they are to be meaningful and valid at all. Reasons, among them moral reasons, are not private mental entities, but relational in nature. This renders them public in the sense that they are shareable even when not shared. The relation can be established with ourselves or with others, and comes in the form of law.

There still seems to be a gap between my reasons and yours, though. Not only must my moral reasons be shareable in the sense that they can be understood, but they must somehow be acknowledged by the other as his reasons, too. Why should I accept your reasons as mine? Or, better, how can I morally obligate you? Korsgaard appeals to an argument from Thomas Nagel (Nagel, 1970): If you are torturing me, and I command you to stop, something changes.

Adding ‘how would you like it if someone did that to you?’, I have given you a reason to stop. I urge you to consider what it would be like to have something like this done to yourself, which, obviously, you would detest. You would think that if this was being done to you, the other would have a reason to stop – and that reason would be the value you place upon yourself as a human being, a value and a reason we share. So, by reminding you that I am a human being, just like you, which it would be hard for you to deny, and that you are a human among equals, in the sense that we all share the same humanity, you would have a moral obligation to value me like you value yourself. So the gap in this situation – between me valuing my humanity and you valuing my humanity – is only there if there is a reflective failure on your part, a failure that I am trying to correct by my proposition that you change places with me. Of course, it could be the case that you assign lesser inherent worth to me or greater inherent worth to yourself, and that you could fail to see that we have anything in common. But it seems you have no reason to do so on Korsgaard’s account, no good reason that can withstand reflection, that is. If I listen to your argument at all – something that I can hardly fail to do if we speak the same language – I have already in some sense acknowledged your humanity (Korsgaard, 1996a, pp. 142-3).

Korsgaard believes that arguments against the privacy of consciousness in the sense just described are arguments against the possibility of egoism. Egoism is defined as the view that only your own interest is the source of reasons. She agrees with Nagel that the egoist is a practical solipsist, and against such a position she argues that ‘you can no more take the reasons of another as mere pressure than you can take the language of another to be mere noise’ (Korsgaard, 1996a, p. 143). So others can obligate us on account of their being human beings at all, and also on account of any communicated reasons, as long as these reasons pass the test of reflection. Your appeal to yourself as a human being with inherent worth is a good and moral reason for me, a reason that I can’t avoid taking into consideration if I am rational. And as a human being, rational is exactly what I am – at least by capacity, if not always in action.

Korsgaard goes on to argue for the power of animals to morally obligate us, through a reference to physical identity, an even more fundamental identity than practical identity. Humans, of course, share this animal nature, and have to value it if they are to value their humanity. Value is implicit in the fact of life, although it takes a valuer, a human being to

acknowledge this. We might say that this is just another implication of our human *form*, in the Aristotelian sense. What Korsgaard is saying is that what we are, our form, is not just our specific capacity for rational thought and action, *logon echon*, we are of course *zoon* as well, which gives rise to the fundamental value of life as both its own end and a precondition for valuing anything else. Continuing the comparison with Aristotle, one may say that the previous discussion of the public character of reasons and the comparison with language might be understood as a way of explicating man as *zoon politikon*. That Korsgaard's view is that reasons and value emerge intersubjectively rather than being strictly agent-relative or agent-neutral is something that emerges even more clearly in an essay from Korsgaard's, *Creating the Kingdom of Ends*, where she attributes inherent shareability to reasons and claims that the existence of reasons relies on the existence of agents (in the plural) (Korsgaard, 1996b, p. 276).

An interesting discussion of pain also emerges in the context of animals and the normative status of pain, in which Korsgaard defines pain (both physical and mental) as the perception of a reason to alter our condition, a perception that could be wrong. This means that human pain must also pass the test of reflection to become a proper reason – after all, there are pains that we would not want to be without, and pains that we have an obligation to endure. This is offered as an alternative to the view that pain is an intrinsically normative entity.

Korsgaard returns to the contingency left in lecture three, the topic of the practical normative sceptic, who denies that anything has value. He would be devoid of reasons to do anything, since reasons and ends require value. The latter follows from what we might call the rationalist definition of intentional action: 'when we make a choice, we must regard its object as good' (Korsgaard, 1996a, p. 122). This is a precondition of rational action that Korsgaard doesn't discuss too much, although the power of her account seems to hinge on the truth of this claim. One might also turn the problem around, and say that Korsgaard's picture of human nature contributes towards the justification of her view of intentional action. I will return to this in later chapters. At any rate, Korsgaard claims that as long as we remain living human beings, we

> *have to* engage in rational action. Animal action, unreflective action, is not open to us; and yet we must do something. So, does the normative sceptic, after all, have to commit suicide? There is no way to put the point that is not paradoxical: yes and no.
>
> (Korsgaard, 1996a, p. 164)

That is, yes, he would have to commit suicide, since there would be no reason to live. But then again, he would have no reason to commit suicide, either. Nor would he have a reason to do anything else. He could follow the desire of the moment, but on Korsgaard's account this would be impossible, since being merely an animal is not an option.

Summing up, Korsgaard claims that the positions of voluntarism and realism turn out to be true, after all. Voluntarism is true in the sense that lawgiving is the source of normativity. Of course, for Korsgaard the authority of the lawgiver is the authority of your own thinking self. This self, in the vein of Freud's superego, also has the ability to punish and motivate you through a guilty conscience, regret, remorse, fear of loss of identity and the like. However, it is not dependent on the negative moral emotions in order to have authority; rather, it needs them to perceive its reasons, and thus to operate at all (Korsgaard, 1996a, p. 151). The role of emotion thus becomes prominent; it is a requirement of the functioning of a rational mind. Further, Korsgaard maintains that realism is true in that there are intrinsically normative entities, the entities for Korsgaard being human beings, in virtue of the power of reflection. Value, it turns out, is both in the world and in the mind (Korsgaard, 1996a, p. 155). The method of reflective endorsement is the way to tell right from wrong; we discover through reflection that autonomy is the source of obligation, and that moral normativity is the result when we discover that our fundamental practical identity is that of a human being, requiring us to grant others the same respect that we afford ourselves. This is an argument for the Humanity formulation of Kant's Categorical Imperative, and possibly implies the moral law, that we should see ourselves as members of the Kingdom of Ends, where everybody's humanity are taken into account.

And so it could seem that Korsgaard has completed her aim; to come up with an account that not only explains, but also justifies morality, in answering the normative question, which can be rewritten: 'why should I be moral?'. If I attempt to answer this briefly in the vein of Korsgaard it will go something like this: I should be moral because I am the sort of being who needs reasons to act and live. The reasons are provided by our practical identities, the roles with which we identify, and those reasons and identities can in turn be further questioned. Among my bundle of contingent identities, however, what is non-contingent is that I must see myself as a human, respecting my ability to value if I am to confer value onto anything at all. This comes with the requirement of respecting that same human capacity in everyone, which

in turn requires of me that I grant them the possibility of exercising their rational powers, being laws to themselves as they deem fit, as long as the laws they then live by are not in violation of this fundamental law, that humanity must be respected. This provides us with at least a basic notion of what morality is, and a test of our principles of action (and a test of our contingent practical identities), in that they can be measured against the standards of universalisation and respect for humanity. The more detailed implications of this, what exactly morality demands of us, would require further determination. Korsgaard does not attempt this here, referring only to an effort she probably respects, that of John Rawls, when it comes to a more detailed decision-procedure for the laws of the Kingdom of Ends.

Returning briefly to Prichard, I think Korsgaard believes she has accomplished an intrinsic justification of morality, as long as reflection recognises that it is fundamentally moral by implication. If you fail to be moral, you fail to live up to the standards of your distinguishing feature as a human, that of rationality. Reflection approves of itself as moral. If it is in our self-interest to be moral, this seems to be a welcome bonus in Korsgaard's case.

2 The argument and its concepts

In the beginning of chapter three I will consider some of the criticisms included in the latter parts of *The Sources of Normativity*, and then I will turn to other sources, including myself, for various challenges to Korsgaard's argument. In this chapter, I will try to make the central parts of that argument clearer by way of analysis, clarification of concepts and finally an illustrational model of deliberation, as I believe Korsgaard conceives it.

2.1 The arguments

2.1.1 The need for reasons

> P1: Human beings have a reflective capability that enables and forces them to question their *incentives*[11].

[11] Korsgaard uses 'desire', 'inclination' and 'impulse' as well. I**ncentive** is the most basic concept. In her reply to her critics in *The Sources of Normativity,* she defines 'incentive' as 'a desire or other impulse that presents a certain action as worth doing' (Korsgaard, 1996a, p. 243); in *Self-Constitution*, she equates it to Kant's concept 'Triebfeder' and defines an incentive as a 'motivationally loaded representation of an object' (Korsgaard, 2009, pp. 104-5) that makes the object 'attractive or aversive from some point of view' (Korsgaard, 2009, p. 120). 'Object' is here used in a wide sense, to include states of affairs and activities. 'Inclination' or 'desire', in Korsgaard's language, is our self-conscious awareness that there is an incentive. This, in turn, faces you with a choice and a demand for a reason. For Korsgaard, as we have seen, 'I want this' is not by itself a reason, it means we are aware that we have a candidate for a reason, a subject of reflective endorsement or rejection. Having an inclination or desire follows from your responsiveness to the highlighted features of the object of the incentive; the desire for pursuit or avoidance follows from pleasurable or aversive features of the incentive. Thus, an incentive is something that suggests action, upon cognition (my term). By cognition, I mean that the suggestion can arise as a *perception* that suggests action, like a *seeing* a person lying unconscious on the ground, or *hearing* someone tell you to do the dishes; on the other hand, the suggestion can arise from *imagination,* like when dreaming or day-dreaming results in an idea that might be worth putting into action; or a sudden recollection from *memory* about something that you were supposed to do. The list is not exhaustive, I only mean to indicate that there are several ways in which an incentive can present itself. At any rate, motivational considerations surface as a part of the incentive; if the representation is attractive, we may find ourselves immediately desiring to pursue it, like hunger suggesting eating. In the case where the representation is aversive, we find ourselves desiring to avoid the object or to do nothing, like when the alarm bell goes off, suggesting we should get up. In the latter case, the desires are mixed. We have of course willed it when we set the alarm clock the night before, and we are obligated unless special circumstances suggest otherwise, on Korsgaard's account. But there is certainly *some* motivation to the contrary; there is the unpleasantness of leaving bed that has to be fought in order to do what we set out to do. One will have to say that the alarm bell going off triggers two (or more) incentives, one to get up and one to stay in bed. Then (if in

In the case where we don't question our incentives, we are still responsible for our action because there is implicit endorsement whenever we carry out an intentional action.

> P2: Questioning our incentives means having to *choose*[12] whether we act on them or not.

> P3: The choice must refer to *reasons for action*[13].

The reflective process that leads from incentive to action is known as practical deliberation. The conclusion of practical deliberation is action, not the formation of an intention (Korsgaard, 2009, p. 124).

> C1: It follows from P1-3 that human beings need reasons in order to act.

Korsgaard thus defines a human being as 'a reflective being who needs reasons to act and live' (Korsgaard, 1996a, p. 121).

2.1.2 The general form of reasons for action

> P4: The incentive suggests an end, something to be realised, and is a candidate for a reason for action.

> P5: *Reflective endorsement*[14] of an incentive on account of a principle of action (maxim) constitutes a reason. Reflective rejection constitutes an obligation.

If there is rejection, it means that we can't endorse the proposed end. This can happen when either the end or the required means are at odds with some other principle of action that we endorse, and that we judge to provide better reasons for action. Ends, then, are chosen as parts of maxims which in turn are chosen as laws the agent gives to herself (Korsgaard, 1998, p.

---

doubt) we have to decide which one is the better candidate for a reason, in light of some principle of action or practical identity.

[12] To **choose** is to follow the principles of practical reason, the hypothetical and categorical imperatives (Korsgaard, 1996a, p. 236). When we act on a choice, we also choose who we are, i.e. our practical identity.

[13] The **reason for action** is 'the incentive as seen from the perspective of the principle of choice' (Korsgaard, 1996a, p. 243)

[14] **Reflective endorsement** or rejection is an acceptance or denial of the incentive's candidacy for being a reason.

55). The laws to which you already subscribe are your *constitution*[15], on account of previous legislation by the free will. You need better reasons in order to overturn those laws. Until that time, you are obligated by the laws you already have.

> P6: It is the law of practical reason that the principle of action must take the form of law and universalisation; this law of practical reason is known as the categorical imperative: when reflectively endorsing a proposed end, we must will always to do *act*[16] φ in order to realise end E in circumstance C.

When we carry out an action, the action becomes an embodiment of a particular law that has the form of the categorical imperative. However, the right form does not establish what the content of the law should be.

2.1.3 Practical normativity is established by practical identity

What, in particular, we should and shouldn't do is a product of previous self-legislation, or if no relevant legislation is internalised, the point of view we choose to take up in deliberation as a basis for new legislation. For Korsgaard, whether the incentive is a reason depends on whether it is recommended by some normative standard to which you subscribe by virtue of your identity, i.e. who you think that you are.

A practical identity is 'a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking' (Korsgaard, 1996a, p. 101). As such, it is ultimately from the point of view of some practical identity that a proposed action is endorsed or rejected as a specific act for the sake of a specific end.

---

[15] **Constitution** is here understood in the dual sense of constituting your identity and laying down laws for yourself. This is a principal idea in Korsgaard's *Self-Constitution*, where she argues that the function of action is constituting yourself well, which corresponds to Plato's conception of a well-constituted soul, likened to a well-constituted republic in *The Republic* (Korsgaard, 2009, p. xii).

[16] An **act**, understood as what is carried out, does not make sense alone, but only in relation to some purpose. An **action** is an act-for-the-sake-of-an-end, and the relation between them must also be described in order to make sense of the action as worthwhile. More on the distinction of acts from actions will follow in section 2.4.

P7: You have practical identities that provide the normative standards – i.e. principles of action, laws, or values[17] – required for endorsement or rejection of an incentive.

A practical identity, I_x, can as far as I can tell be understood as a set of particular values, such that $I_x = \{V_1, V_2, \dots, V_n\}$, where V_y is some particular value and n is the number of values contained in the identity. Values come in the form of particular commands of the general form of the categorical imperative. One can imagine that these values don't have to be spelled out, only that they can be, and that they are derivable from I_x through grasping the normative standard inherent to it. In general, the set model I here extract from Korsgaard's views is intended as a tool for understanding what she is getting at, not to be taken too literally, and certainly not to the effect that such considerations are always explicit to the deliberating agent.

Your total identity, TI, what you identify as overall, will be the superset of practical identities $TI = \{I_1, I_2, \dots, I_m\}$, where m is the total number of practical identities, i.e. the total number of descriptions under which you value yourself. TI will typically be filled with I's, looking like this: stock broker, Barcelona Football Club supporter, Catholic, husband, cat owner, Catalan, etc.; perhaps there will be more general notions: good person, member of the Kingdom of Ends; perhaps there will be more specific ones: dandelion lover, porcelain figurine collector. By 'general notions' I mean that some identities seem to demand a say in more, if not all of what you are doing, while others appear to have a more specific domain. The more general an identity, the more compatible it should be with the overall system, if coherency, avoidance of conflict between identities, is a goal; for in the total identity TI, it is entirely possible that incompatible values exist; if not in the same set I_x, then across two such sets to which we subscribe. To the extent that our aim is a coherent set, any values that are incompatible should be brought into harmony either by abolishing one of the particular values V_y or the entire set I_x to which it belongs, if I_x requires V_y absolutely and V_y violates other values that are of greater importance to us by virtue of the identity they belong to.

It seems, though, that not all the descriptions under which we identify are descriptions under which we value ourselves. Addiction is perhaps the most obvious example; you are a cola addict, you identify as a cola addict, but you do not value yourself under that description. Yet

[17] As far as I can tell, the three are equivalent or near-equivalent. **Values** may be slightly more general and emotionally loaded, though probably not for Korsgaard who assumes that the relevant motivation issues from rational conclusions.

you can't help acting like one, and it does come with the normative standard of drinking Coke or Pepsi whenever you feel like it, which in your considered view is far too often. There may be a general problem here, not only with addiction, but with weakness of will, acting against one's better judgment. This a thought to which I'll return; at any rate, the identity of a cola addict doesn't qualify as a practical identity according to the definition, since it is not an identity under which you value yourself.

2.1.4 You must value yourself

There are consequences of the definition of practical identity that may be listed as premises to a further conclusion:

P8: You have a practical identity iff you have reasons.

P9: You have a practical identity iff you can be said to value yourself.

C2: A person has reasons iff he values himself.

In other words, you *must* value yourself (see rational necessity below), on pain of having no reasons, reasons you need in order to live and act.

Certainly, the opposite seems to be a possibility, the most obvious case being an all-encompassing depression, in which case not valuing yourself and having no reason to live or act are symptomatic. The meaningfulness of the world and the meaningfulness of possible actions withdraw during depression. Korsgaard does not talk about depression, but considers a philosophical stance to the same effect, the case of complete normative scepticism I referred to in chapter one (Korsgaard, 1996a, pp. 160-4). At any rate, it is hard to imagine someone completely bereft of reasons; a person who commits suicide, for instance, needs to reflectively endorse their desire to do so, which on the present account would mean that they value themselves as a source of reasons, which means that they have a reason to sustain life. Seeing oneself as completely without reasons would mean complete apathy, which could lead to a passive form of suicide from lack of sustaining life. But one would be hard pressed to find a real-life example like *that*. What this seems to show is that suicide is irrational. Other than that, one might have a rational reason to commit suicide, or sacrifice one's life for the sake of humanity in extreme conditions.

What does the value of valuing oneself look like? As stated above, I understand values, on Korsgaard's account, to come in the form of laws. The law or principle of action associated with valuing oneself, then, must be something like Kant's Humanity Formula, just first-personally: You must through your actions treat your own person as an end, and not just as means.

2.1.5 You must value others

To get to the third person, some further argument is required:

> P10: The sense in which you value yourself is that you value your humanity, i.e. your ability to confer value by willing in accordance with the categorical imperative.

Note that in Korsgaard's system, all action is autonomous; Kant's delineation of the principle of self-love vs. the Categorical Imperative is replaced by contingent identities and necessary moral identity, both sources of autonomous action (Korsgaard, 1996a, p. 243), but only the moral identity is a source of moral action. Valuing yourself as the source of your values and reasons means valuing your nature as a value-conferrer; but are you not just valuing *yourself* as the source of the particular values you happen to have, and not your *nature*? Do you have to value that you have the ability to value? Well, without valuing that ability, at least, your other values could not subside. So you must value what makes them possible. Thus, what Korsgaard wants you to value most deeply about yourself is the condition of possibility [Möglichkeitsbedingung] of rational action; that value-conferring ability, i.e. your reflective ability, which is shared by all. In other words, the transcendental argument aims to show us that we must value humanity as the condition of possibility for rational action.

This does seem abstract. But what do we usually mean when we accuse someone of inhuman action, or crimes against humanity? I would say we are accusing them of trampling on the value we place on humanity, perhaps on life itself in the case of attributing 'inhuman' to, say, the torture of animals. The fact that human lives matter, that they are worth something and that actions that suggest otherwise are atrocious to us, can be explained in a number of ways, which is the subject of ethical discussion. Korsgaard's way of explaining the value of others as a rational commitment is indeed abstract, and reference to a natural feeling of repulsion, like in Hume, speaks to us in a more immediate way. The strength of Korsgaard's way of approaching the justification of morality is not its immediate availability; rather, she gives a

voice to the possibility of such condemnation being warranted from a rational point of view, and subject to something more than biology or whim.

Your values might change, but your ability to confer them remains. And as such, if you come to value your humanity, what you value is not strictly subjective, but intersubjective[18], common to all men. It follows that you value something about others, because it is identical to what you value about yourself; in a word, on pain of irrationality, you must value the humanity of others.

> P11: Since there is no difference between your humanity and that of others, and since reasons are essentially public, you must necessarily value their humanity.
>
> C3: It follows from C1, C2 and P10-11 that insofar as you have or are to have any reason to live and act at all, you must value the humanity of others.

Identifying under such a description (one who values the humanity of others) is a necessary practical identity: *moral* identity. Again, this is contingent on you valuing anything at all. It is also a description of the ideal, rational agent, something we should keep in mind when presented with a strong formulation like C3. Imperfect rationality and contingent identities can be said to provide reasons on their own, but they are not the best, most justified reasons.

Returning briefly to the thought of TI as the total set of identities, one could say that moral identity, if adopted, becomes the organising principle of this otherwise ungrounded set. Even without the moral identity, the goal of a coherent set could be argued for in terms of self-interest. It seems very plausible that one would be better off having integrity and a consistent set of principles of action, with regard to issues of conscience and the general taxation that making sense of oneself places on the human psyche. Even so, such a set would not have any restrictions on what the principles should look like, only that they be coherent. The moral identity, however, places restrictions on what the set could look like for the rational agent; if Korsgaard is right, the rest of the set must conform to it on pain of irrationality. If coherency leads to psychological well-being, that is certainly a plus, but not the decisive factor in this account.

---

[18] Korsgaard's general view of values is what she calls an **intersubjectivist** view; that they exist for all rational agents, but not independently of those agents. Objective values are constructed from subjective ones, the reverse being the case in a substantive realism about values. See *The Reasons We Can Share* in *Creating the Kingdom of Ends* (Korsgaard, 1996b, pp. 278-9).

When aiming at right action, then, it turns out that in addition to following the principles of practical reason, the structure of reason itself commits you to having values, valuing yourself and valuing others.

The argument shouldn't be understood as suggesting the dubious notion that you already and explicitly value yourself on account of your ability to value. What Korsgaard is saying is that we have a rational commitment to value our humanity insofar as we have, or are to have any values. We thus have a reason to adopt the moral identity as an explicit resource, a source of authoritative reasons. Moral autonomy becomes the final, rational end of action.

Now, I want to briefly supply this view with the one Korsgaard presents in a more recent work than *The Sources of Normativity*.

2.1.6 Agency and self-constitution

A slightly different take on why our total identity should be ruled by reason as well as be coherent, is offered in Korsgaard's *Self-Constitution*:

> P12: The function [constitutive aim] of action is self-constitution
>
> (Korsgaard, 2009, p. xii)

This means that the goal of action and life, conceived as the *form*, in the Aristotelian sense, of action, is to constitute oneself well. Being well-constituted means having a coherent set of laws – a constitution – self-legislated on account of moral identity as fundamental.

A point of this formulation is to reinstate some of the force of our contingent identities by way of saying that what we *are* is not just our rational capacity, i.e. a reason fighting the alien passions[19]; we are our constitution, i.e. we are what holds us together in a republic of the soul ruled by law, where passions are welcome and necessary but subject to that law. In Plato's *Republic*, like in Korsgaard's view, the constitution best suited to achieve an integrity of the soul is that in which reason is made the lawful ruler. Constitution based on other principles will undermine any integrity or coherency of the soul. So in one sense we *are* our desires as well, but these would tear the single, unified will apart if they took turns ruling.

[19] A view of which Kant has often been accused. Korsgaard takes Kant's considered view about desires to be less severe than this (Korsgaard, 2009, p. 154).

2.2 Rational necessity

The necessity that arises from the arguments as outlined in 2.1, like that of valuing the humanity in others, is *rational necessity*. It is a first-personal necessity; a necessity that the rational agent *confronts* in deliberation (Korsgaard, Normativity, Necessity and the Synthetic a priori: A response to Derek Parfit, 2003, p. 7). However, all agents share a rational capacity, and in this sense, the necessity is universal – it is experienced by anyone who exercises their rational capacity. The principles of *practical* reason are the hypothetical and categorical imperatives, whose rational necessity surfaces in deliberation. We can fail to acknowledge them, but we *should* acknowledge them*, if we are rational*. Thus, rational necessity can be written as a 'should'. Korsgaard likes to exemplify this by something more familiar, the principles of *theoretical* reason: if you believe the premises, you *should* believe the conclusion. That is, you *must* believe it if you are guided by reason (Korsgaard, The Sources of Normativity, 1996a, p. 226fn). But you don't *have to*, in the sense that you cannot fail to do so. Likewise, if you will the end, you *should* will the means, i.e. you *must*, if guided by reason; and if you will an end, your willing it likewise *should* have the form of universal law, or *must* if you are guided by reason. Note that this rationality clause, 'if guided by reason' is essential to the idea of rational necessity, an idea that figures at the heart of Korsgaard's argument. It is in the reflective perspective that moral normativity arises by rational necessity. This notion of necessity, as we have seen, is contingent on the agent being rational, which he may fail to be for a number of reasons. The notion of rational necessity implies that there is a rational *ideal* that we, with our reflective capability, can aspire to reach. Korsgaard is often assumed to believe that all intentional human action is rational, all the time, which makes a mystery of all the examples to the contrary. Often, when Korsgaard writes a sentence such as 'you must value yourself', the rationality clause is omitted, yet implicit. The notion of rational necessity thus turns out to be a considerable source of confusion, as we will see when dealing with critics of Korsgaard in chapter three.

2.3 Internalist views in Korsgaard

Drawing on Bernard Williams, I now want to get back to some tough philosophical issues, briefly elucidating Korsgaard's position on two different species of internalism and externalism.

Reasons internalism

Like Korsgaard, Bernard Williams is an *internalist about reasons*. This means that whatever an agent has a reason to do must correspond to the agent's *subjective motivational set*; you cannot have a reason to do something for which there is no basis in that set. Externalism about reasons, then, implies that no such motive is necessary. For the externalist, you can be said to have a reason even if it is impossible for you to act on it. Williams' claim in his classic paper *Internal and External Reasons* (reprinted in *Moral Luck,* Williams, 1981) is that the notion of external reasons simply make no sense, as the notion that there is no motivation present doesn't make sense of the statement 'A has a reason to φ'. We can't explain his actions without reference to motivation. Williams writes:

> The whole point of external reason statements is that they can be true independently of the agent's motivations. But nothing can explain an agent's (intentional) actions except something that motivates him so to act.
>
> (Williams, 1981, p. 107)

Williams also includes the process of deliberation as something that can alter your existing motives, but not independently of what was already in your subjective motivational set. The set itself may be altered on account of deliberation, but not in a way that spawns motivation independent of what was already in the set before deliberation. In Williams' words: 'we should not, then, think of S [the subjective motivational set] as statically given'. And though one can see the set as consisting of desires, 'this terminology may make one forget that S can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties and various projects, as they may be abstractly called, embodying commitments of the agent' (Williams, 1981, p. 105). It seems fairly clear that this is compatible with Korsgaard's concept of practical identity.

In accepting Williams' internalist account, Korsgaard further has to assume that all humans, by virtue of their capacity for rational thinking, already have a motive – sometimes slumbering or implicit – to do the rational thing, a motive awaiting the relevant deliberative process in order to be activated. And once the right procedure of deliberation is complete, one will be motivated to do what one concludes is the right thing. This leads us to another position:

Motivational judgment internalism

Korsgaard's position is also a form of motivational judgment internalism – the view that we are always motivated to some extent by what we conclude is the right thing to do. If agents, as they often do, fail to do what they have considered to be the right thing, there is some sort of deliberative failure, a failure of practical reason. On the other hand, every agent, with his capacity for rationality, must at least have the *capacity* to be motivated by such considerations as may arise from rational deliberation. Since Korsgaard employs the concept of practical identity to account for all the non-moral, normative commitments in our lives, this has to hold for normative judgments in general, not only moral judgments: 'in some cases our conception of a contingent practical identity will give rise to new motives in a way that parallels the generation of the motive of duty by the thought of the categorical imperative.' (Korsgaard, 1996a, p. 239). A Korsgaardian agent, when deliberating about what to do, will consult his practical identities that give rise to normative obligations that he may or may not act on. This is Korsgaard's way of explaining how the agent's subjective motivational set relates to practical reason. In a sense, both Williams and Korsgaard see *who we are* as giving rise to what we have reason to do. Korsgaard, of course, has to make additional assumptions about the moral reasons being the crown of rationality, so to speak, guiding other rational considerations when the fully rational agent comes to see and affirm their authoritative place in his subjective motivational set.

The relation between reasons internalism and motivational judgment internalism is a close one, and seems to be as follows:

> Reasons Internalism: If something is to be taken as a reason by an agent, the agent must be able to be motivated by it

Williams requires that the agent must be able to be motivated if we are to be able to explain his action and render it rational; motivational judgment internalism in its general form adds that agents are motivated by their judgments about what their reasons are;

> Motivational Judgment Internalism: What an agent judges to be normative reasons always motivates him to some extent

Here, we might add 'insofar as he is rational', and still be on speaking terms with Williams, if I understand him correctly; Korsgaard, however, thinks that rationality has further implications than instrumentalism. Her account of the ideal reflective process is one that

renders some ends final. If *moral* reasons are to have any force, you have to be able to be motivated by *them*.

> moral rationalism: if something is morally wrong then there must be a reason not to do it
>
> (Finlay & Schroeder, 2008)

This is a general formulation from the *Stanford Encyclopedia of Philosophy* entry on reasons for action, and the possibility of not having a moral reason is still available on such terms. If you are not motivated, you don't have a reason. But on Korsgaard's terms, you do. This is because her reasons internalism is a *counterfactual* version of reasons internalism, in claiming that the agent *would* be motivated, if he were rational; this as opposed to an *actual* version, where the agent must actually be motivated if one is to say that he has a reason (Finlay & Schroeder, 2008). Taken together with the view that all agents have the capacity to be rational, I can extract from Korsgaard an argument to reach the conclusion that moral reasons are universal in the sense that they apply to all agents, counterfactually. It would look something like this:

> Everyone would be motivated by moral reasons, if they were rational
>
> Everyone has the capacity to be rational
> ---
> Conclusion: Everyone could be motivated by moral reasons

> Moral reasons apply to everyone who could be motivated by them
>
> Everyone could be motivated by moral reasons
> ---
> Conclusion: Moral reasons apply to everyone

So moral reasons apply to everyone because all agents have the capacity to be motivated by such reasons. This means that if agents are not so motivated, they should have reflected more, which would have motivated them.

Finally, I take note of the fact that Korsgaard's own formulation of what she calls the 'internalism requirement' in *Scepticism about Practical Reason* leaves the door open for scepticism:

> The Internalism Requirement: Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons
>
> (Korsgaard, 1996b, p. 317)

This is not the claim that such reasons *do* in fact motivate any action, for that can't really be known. It is the claim that such reasons have to be internal if they are to have any relevant existence.

2.4 Justified reasons for action

Now, we turn to something slightly different, but related. The kind of reasons I am talking about when discussing internalist views are primarily normative or justified reasons. From discussing the requirement that reasons be able to motivate, I will now outline Korsgaard's picture of how you are justified in acting on your motives, and what motives provide justification. In other words, whether your reasons and your actions are justified relates to the kind of motive you can adduce for performing a given action.

In *Self-Constitution*, Korsgaard provides an illuminating discussion where she distinguishes 'acts' from actions; the act is the manifest part of the action, which cannot be evaluated apart from the action as a whole. The action as a whole is conceived as including the end, so that an action is an act-for-the-sake-of-an-end, both 'the object of choice and the bearer of moral value' (Korsgaard, 2009, p. 12). Thus, it is the action as a whole that has the possible quality of rightness or goodness, and not the act. This is a way of clarifying the confusion about reasons that can arise from a simple question like 'why did Jack go to Chicago?' (Korsgaard, 2009, pp. 12-4). An answer like 'to visit his mother who lives there', might signify that the action is going to Chicago and that the purpose is the reason, but these ideas confuse things, according to Korsgaard; we can only make sense of the action as going-to-Chicago-for-the-sake-of-visiting-one's-mother. So when asking 'why', already knowing the act, we are not inquiring about the reason, because the real reason is in a sense always the same for every action: because it was deemed right or worthwhile by the agent. We ask 'why', inquiring about a purpose, which in turn will cast light on whether we have a good reason or not. A purpose, then, is not by itself a reason, but the purpose can make the action – conceived as a whole – seem worthwhile to us. I might add that even then, it will not necessarily be clear whether it is a good reason; we may have to explicate further if the standard, i.e. the practical identity in question is not an obvious one, unlike the case here about the good son; justified actions for a good son is something most people can relate to. The boy's motive can here be cast as the motive of duty, or more specifically the motives of a dutiful son, which should be endorsable from the point of view of moral identity as well. Practical identities as well as

moral identity can on Korsgaard's account motivate the agent, as referred to in the previous section.

In order for what you take as a reason to be a justified reason, the standard against which the reason is measured must be incorporated in the process of taking-as-a-reason. If your action is not expressive of that standard, it is not justified. And whether such an action were ever realised is impossible to say, given the insecurity about people's motives. That they are motivated by reference to *some* standard for every action, is clear, but what is not clear is whether, when asked, they can know what their real motivation was. For instance, if you believe your action to be justified by the standard of magnanimity, the act itself will look exactly like the one justified by the standard of selfishness: giving alms to the poor in order to alleviate their troubles and giving alms to the poor in order to be more electable as a candidate for political office both share the same *act*, but are not the same *action*. Nor can the action be justified if it is based on false belief about the means to your ends, such as when your end is famine relief, justified in the light of the standard of charity, and you send money to a corrupt regime that uses the money on building a golden palace to honour the despot in charge – even if your end happens to be realised somehow. But your action can certainly be *explained* by reference to such false beliefs. All justified internal reasons must satisfy the demands of explaining why the agent did what he did. But as we have just seen, not all reasons that explain why something is done are justified, even if they remain rational relative to a false belief (see Williams, 1981, p. 103).

When we say that Stalin had no reason to send his opposition to the Gulag, that is of course both wrong and right, given the massive ambiguity of the term 'reason'. But as has hopefully been made clear, it is mostly wrong, unless by 'no reason' we mean that he had other, better reasons that should have trumped the one he acted on. He had an internal reason in that he successfully achieved his ends by the means taken, and his actions are readily explainable in such terms. They are *justified* from the point of view of self-interest, as long as he did what was in his best interest, and justified from the point of view of a practical identity like 'leader of the communist Soviet' – one who believes to know what is best for his people, but is surrounded by people who threaten the system; thus, to protect the system, he acted like he should, according to his interpretation of the normative standard of the good communist leader.

But if we, with Korsgaard, are to say that he had no reason, or better reasons for doing something else, we must do so on account of some other standard than the one according to which he successfully acted, some standard that should have been more important to *him*. And according to Williams, that has to mean that he had some ends that were more important to him, the realisation of which was impeded by his actions. These could very well be moral on Williams' account, or something else, but if he could not possibly be motivated so to act, there could be no justified reason for him to act accordingly.

Let's say that he had justified moral reasons for doing something else. Then, what we are saying is either that there are considerations he should have taken into account, given his own motivational set, or we are saying something meaningless: that he had an external reason to do so, one that need not be able to motivate him accordingly, but that nevertheless somehow was a reason *for* him. So Korsgaard needs to show that given Stalin's rational capacity, he failed somehow by not being most motivated by moral considerations, considerations like respect for the life and liberty of those he sent to labour camps. His mistake is not a flawed deliberation on account of the standards he carried with him, his total identity (TI), if you will; that TI didn't include a moral identity as a regulatory device, at least not in the sense proposed by Korsgaard. His mistake, then, is failing to have constituted himself – i.e. his priorities, his standards, his TI – in accordance with Korsgaard's account of reason, from which morality follows as necessarily as taking the means to your ends, because that is what reflection on our humanity should engender. In addition to asking himself, implicitly or explicitly, 'what are my ends?' or 'what is my practical identity?', Stalin should have gone further and asked himself 'what should my ends be?'. Failing to want to take the means to your ends, unless those means are unattainable or in conflict with other, more important ends, is irrational on both Williams' and Korsgaard's accounts. It is a failure of rationality. Korsgaard then adds the claim that there are ends that it is irrational not to have; the end of humanity, or morality. So she could argue something like this:

> Stalin, like any other human being, had the capacity for rationality
>
> The capacity for rationality entails being able to be motivated by rational considerations
> ---
> Conclusion: Stalin was able to be motivated by rational considerations
>
> Rational considerations prescribe morality, i.e. respect for humanity
>
> Counterfactual: If Stalin had employed his rational capacity to the full (which he did not), he would have come to see that he had moral reasons

From C3 (in 2.1) we get that: Moral reasons are authoritative for any agent with reasons at all
---
Conclusion: From the (authoritative) point of view of morality, a point of view internal to Stalin, he did not have a justified reason to deprive others of their freedom

We see by this argument, or something like it, that Stalin can be said to have had an internal reason not to commit atrocities, but that this depends on a counterfactual argument, that he should have been motivated in accordance with motivational factors that he was capable of, but that were not immediately available to him. The internal reason of morality is thus contingent on the agent reasoning his way to morality.

It is not that hard to agree with Williams that when we act-for-the-sake-of-ends, we have a justified reason only if the action is cast as motivated in accordance with elements that have to be present in our subjective motivational set as inclinations, or made to come alive by reflection on this set, revising the set. Where Williams and Korsgaard part is at Korsgaard's notion of rational necessity, that you are less than fully rational by not acknowledging that in order to have reasons at all, other people's humanity must be a source of authoritative reasons for everyone. Every rational agent has the capacity to be motivated by rational considerations; so since everyone has a rational capacity, there is a slumbering morality in everyone. The notion of justified reasons we then end up with is as follows: If you are fully rational, you have justified moral reasons that are authoritiative for you. If you are not fully rational, the fault lies with you for failing to exercise the capacity on which you rely when making any choice and acting at all. Exercising that capacity would have made moral motivation, motivation from duty, come alive.

So in one weaker sense, you have justified reasons to do everything that is relative to the values embodied in your *explicit* subjective motivational set (or practical identity). But in another sense, the only *truly* justified reasons are moral reasons, or other reasons that do not conflict with these. This is because *implicit* in your subjective motivational set there is a moral identity that should be made explicit by reflection. This duality of justified reasons will hopefully be made clearer by my illustration of deliberation in 2.5.

As far as I can see, Korsgaard's internalism about reasons mainly stays true to Williams but is far more ambitious in maintaining that there is a process of reflection that can make the 'required elements come alive' (Williams in Korsgaard, 1996a, p. 216) in any subjective motivational set. The 'required elements' are of course the motivational elements necessary

for moral action. Williams, for his part, seems to maintain the more sober view that either they are there, or they cannot be awoken, at least not in everyone.

2.5 Korsgaard's picture of deliberation

Some further conceptual clarification seems in order. Korsgaard uses several terms that are technical and fairly well defined, like 'the normative question', 'practical identity', 'the categorical imperative' and soforth. What proves to be more confusing are terms that one might say are under construction. Words like 'moral', 'reason', 'obligation' and 'action' are general terms that are construed in different ways by different philosophers, reflecting different stances on a number of philosophical issues. When it comes to those words, whatever definition we can extract from Korsgaard's theory will be a normative one. Her theory is normative with regards to the use of such concepts, and normative with regard to how we should justify our actions.

An element of confusion arises from the fact that words are sometimes used specifically in Korsgaard's sense, at other times – especially by others – in conflicting senses, sometimes without proper explanation of the implication of the use.

The goal for Korsgaard is the rationalist project of describing *proper* or *full-fledged* action and *proper* reasons in a way that supports a universalist ethics of duty, i.e. a well-founded morality with real obligations. Korsgaard's account of morality thus naturally relates to a discussion of agency, as we have already seen. A principal aim of the philosophy of action is to show what, if anything, separates proper or full-fledged action from whatever else one calls action.

To show how 'moral', 'reasons', 'obligation' and 'action' relate to one another, the process of deliberation is a fruitful starting-point. I have already described it in a number of ways, but now I wish to present it schematically. For Korsgaard, a proposed action in the form of an incentive to do something can be followed by

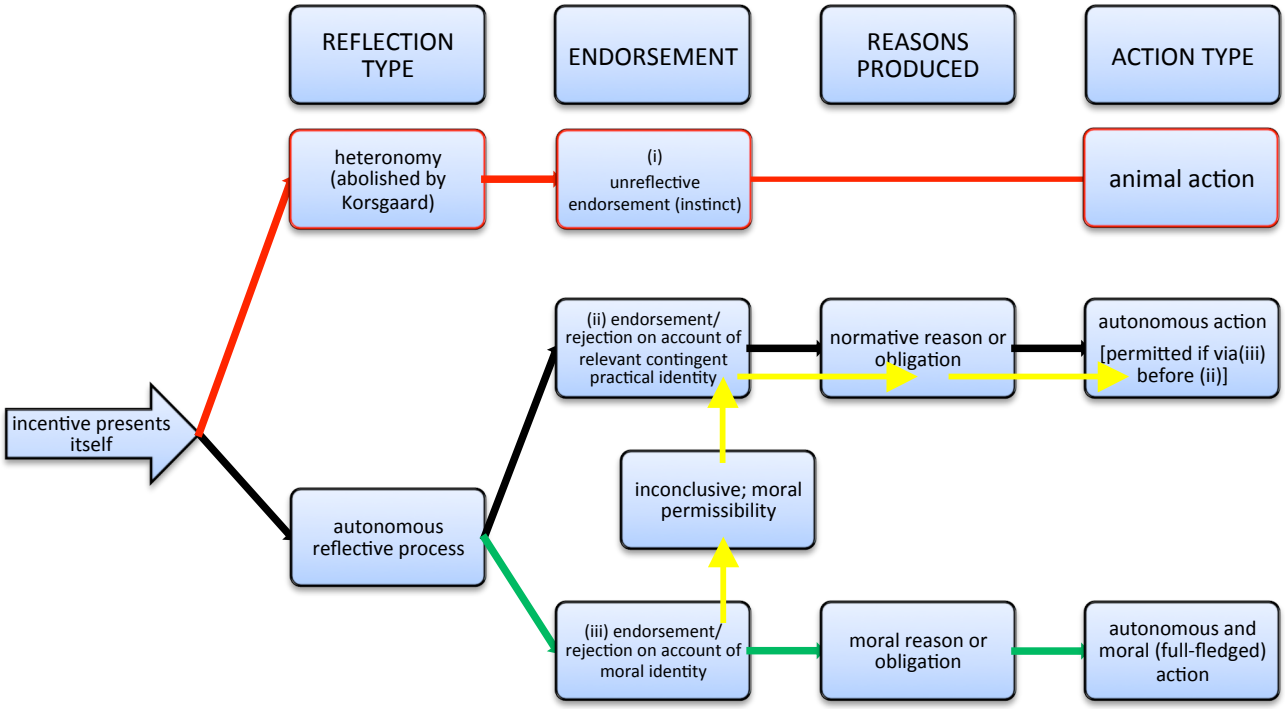
i) Unreflective endorsement, i.e. going with the incentive no questions asked, resulting in action in a lower sense; Korsgaard claims this is only possible for

an animal, by way of endorsement by a principle of action provided by instinct[20].

ii) Reflective endorsement or rejection on account of any of one's practical identities. If reflectively endorsed, such endorsement establishes a normative reason to carry the action through; if rejected, an obligation not to act will be the result. Here, reasons are somewhat justified, in that they are reflectively endorsed, but still not the best available reasons, and not (properly) moral.

iii) Endorsement or rejection in the full light of reflection, i.e. on account of the one necessary practical identity, our *moral* identity as a human being. The moral identity provides a moral reason to carry on or a moral obligation not to, and a corresponding action will be full-fledged.

This yields a basic idea of what Korsgaard thinks practical deliberation, i.e. the reflective process leading to action, should look like. In an attempt to illustrate the idea even more clearly, I have drawn up the following flow chart:



[20] This view is elaborated upon in *Self-Constitution* (Korsgaard, 2009, p. 109 ff.).

Korsgaard abolishes the notion of heteronomy except for the case of animals, as illustrated by the red arrows (i). So even though we may act somewhat like animals if going with our immediate inclinations is our guiding principle, we have nevertheless chosen to do so, which on Korsgaard's account means that we are autonomous. Though reflective endorsement is sufficient for this minimal sense of autonomy, it is not enough to make an action right:

> in one sense no human action can happen without reflective endorsement. When people skip reflection or stop too soon, that is a kind of endorsement, for it implies that the work of reflection is done
> (Korsgaard, 1996a, p. 161)

Heteronomy, then, is not open to us because we endorse our actions by default, even when reflection is incomplete or lacking. As humans, we cannot escape a minimal sense of choice and endorsement, which in turn means that Korsgaard retains the possibility of always holding people responsible for their actions. So what has happened when someone fails to reflect? It certainly looks like they haven't really followed the black arrows, rather the red ones. How can we see this as endorsement on account of a contingent practical identity? A minimal endorsement of one's natural incentive, thus identifying with one's animal nature, but not in the way an animal does, because as a human, one has a choice. As such, one will value oneself under the contingent practical identity of an animal, whether so conceived or not – more likely, you have made a minimal reflection with the implicit conclusion that you always, or at least in this case, have a reason to do what you want to do.

The black arrows (ii), then, show the typical course of deliberation, where some unKorsgaardian notion of morality may or may not play a part, depending on whether it is native to the contingent practical identity or identities one refers to. The identity or identities invoked have to be relevant in the sense that they must have a say about the proposed action, by way of recommendation or prohibition. In the case of no identity having anything to say, it seems you are operating on the fringe of your total identity. Let's say you have never gone hiking, but a friend describes it in ways that makes it seem possibly pleasurable to you; there will be fresh air, magnificent views, marshmallows by the campfire and so forth; let's say you are young or otherwise inexperienced, enough so that you can't say whether this activity is *you* or not. So should you go or not? You don't have the resources to decide, i.e. you have no self-conception that excludes or recommends hiking, only a vague attraction in the form of a favourable assessment by your friend. So you go, and perhaps you find that it was worth your

while, enough so that you now identify as an ‘outdoorsman’ or something to the same effect, which should provide an incentive to go whenever a suitable occasion arises. You have gained a normative conception of yourself, something that carries some weight when deliberation takes place, and that you must make to fit into your total identity.

As for the yellow arrows (ii by way of iii), they depict the deliberative course of the morally permissible. In the case of the morally permissible, so deemed by the moral identity, one is left without a reason to do the proposed action, because it is neither recommended nor prohibited by morality. In one sense, you already have a reason, because you have some inclination, a candidate for a reason, which morality has approved of. But it may still be the case that there are *other* values that speak to the contrary, even if they are not morally relevant. In that case, one will have to refer back to contingent practical identities in order to check whether the suggested action is consistent with one’s optional values. If approved by morality and no other practical identity has any objections, this qualifies as full-fledged action. The reason I believe it is not sufficient to have the go-ahead of morality is that I believe it is hard to imagine an agent that is *only* a moral agent, i.e. where moral identity is his only identity, and that identity getting both the first and the final word about what he should do.

The green arrows (iii) depict the course of deliberation that leads to morally autonomous action. Even though we may label the morally permissible morally autonomous as well, the green arrows signify that morality has a say. If you act on a motive of duty or deny an immoral impulse, then, you retain full rationality and moral autonomy.

Now that we are about to encounter differing opinions, it will serve us well to keep in mind that implicit disagreement as to the implications of different concepts and ideas may lead to misunderstanding. This seems especially to be the case with the idea of rational necessity. But first, I want to expand a little on the general concept of practical identity, both because it is essential to Korsgaard’s argument and because it is interesting in its own right.

2.6 A psychological take on *practical identity*

When an incentive presents itself, it is already a representation of an object in a favourable or unfavourable light. This light, the motivation, must come from something that is *you* in some wider sense than a practical identity. We usually call them inclinations; a tendency to represent some kind of object (in a wide sense) favourably or unfavourably. In Korsgaard's language, an inclination is an object of self-consciousness, something that arises as a consequence of our reflective ability; it is the consciousness that you are attracted to acting in the way that your incentive suggests (Korsgaard, Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind, 1998, p. 51). As such, it is specifically the inclination that is the object of deliberation. All inclinations are in some sense you, even though you might not identify with any conception of yourself that evaluates them positively. So built into the incentive is an unreflective rejection or endorsement on account of something, something that might fit into a description of what one is, regardless of whether one so identifies. You may be more inclined towards fighting and confrontation than flight on account of you being a biological male[21], having such and such hormones and so forth, but evaluate the inclinations negatively in most settings, and if positively, not on account of valuing yourself as a man, at least not in a sense that couldn't be shared by women[22]. You may have what one might call artificial inclinations, a product of culture rather than nature, like being negatively inclined towards people of a different skin tone on account of an upbringing by racist parents, while at the same time not valuing yourself under such a description. You may be a Mafioso with a natural inclination to care about the feelings of others, while not valuing yourself under any conception that allows such inclination. We have many such inclinations of physiological or psychological origin, most often without being able to explain or conceptualise them beyond the language of pleasure and discomfort; I find myself wanting to fight; I find Hispanics unpleasant; I find myself not wanting to shoot the man who didn't pay his protection money. I may have self-knowledge enough that I see myself as racist, and in a sense so identify, while legislating to the contrary. If that is the case, I will to treat others fairly regardless of skin tone, but cannot help myself; I fail to do what I

[21] I doubt there are people who value themselves under those exact words ('biological male'), but if 'alpha dog' or 'man's man' can be taken to have the same type of ramifications, surely there are such people, who would take pride in fighting when it is in no way necessary or justified.

[22] An example of a better description of what one might reasonably evaluate positively in a 'man' is the virtue of courage, as in 'courageous' as the conception under which one identifies.

will. In other words, all incentives have to rely on a negative or positive inclination of yours, and in that sense it is all you. You may not value yourself under some description that is nevertheless a part of you – which is why you want to fight it. But until you win that fight, your will is in a sense unfree. This corresponds to Frankfurt's notion of freedom of the will:

> It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will. And it is in the discrepancy between his will and his second-order volitions, or in his awareness that their coincidence is not his own doing but only a happy chance, that a person who does not have this freedom feels its lack. The unwilling addict's will is not free.
>
> (Frankfurt, 1997, pp. 20-1)

Frankfurt's claim is that you have to satisfy your second-order volitions – you must act on the desires that you decide that you want to act on – in order to feel like a person at all, one who is not estranged from himself. This corresponds to Korsgaard's agent who acts or fails to act on the prescriptions of his practical identities; here, it is not a question of morality, only a question of acheiving coherency or integrity. You are not free if you fail miserably at acting on the desires that you have decided that you want to act on. You fail at being what you would like to see yourself as. In that sense, practical identities is not so much who we are as who we would like to be, in most cases. The rational agent, of course, brings harmony between the two.

In notions borrowed from the psychology of Carl Rogers, one could say that you lack integrity if your *perceived* self, what you see yourself as doing, diverges too much from your *ideal* self, what you think you should be doing (Rogers, 1961). Your ideal self, then, corresponds to your second-order volitions or practical identity, while your perceived self is how you see yourself as actually acting. Things are further complicated by the fact that your perceived self may not be accurate, it may diverge from your *real* self – what you are actually like, as opposed to what you believe you are like – by way of self-deception, a fact noted by Frankfurt as well (Frankfurt, 1997, p. 21). The alcoholic may not admit that he is an alcoholic, not even to himself; perhaps this is because alcoholic is not a part of his ideal self or practical identity, and because it is then a load off his mind to *perceive* himself as not being an alcoholic. Facing his *real* self would face him with a problem that his self-deception has buried. That our mind may very well be able to accommodate the breaking of laws while preserving well-being to some extent is empirically well documented; psychological defense mechanisms like rationalisation – the reconstruction of our behaviour as rational when it is not – as well as cognitive biases, like selective perception, may suggest that there is empirical

support for Korsgaard's claim that we are the kind of beings who strive for good reasons and rationality to guide us. When we fail, we try to evade the ensuing pain by recasting ourselves in a better light. Rogers, operating from a psychological point of view, would have it that therapy should bring us to become more honest with ourselves – make the perceived self congruent with our real self – and finally that those be congruent with our ideal self in a process of *self-actualization*. This seems fairly close to Korsgaard's ideal of integrity, though she probably[23] has stricter demands on moral guidance, and the direction this assimilation should take, which means that on her view, you should aspire to become rational, not lower the bar by adapting your ideal self to match whatever it is that you are to feel better about yourself.

All this is to say that our self-image is more complicated than just a reference to practical identity, which is in one sense the part about ourselves that we value, our ideal self. We have conceptions about ourselves that we don't value, like that of the cola addict, but this is not a part of our ideal self, so we strive to get rid of this unwanted, perceived self. But we can value ourselves as, say, a teacher, without believing that we are being the best teacher that we can be. And so we strive to become that ideal teacher. Like Korsgaard says, our practical identities can withstand a few blows. But we also see that practical identity is a psychological concept that makes sense as a carrier of the standard that we try and live up to, the normative standard that we have legislated for ourselves.

[23] I don't know enough about Rogers' psychology to account for whether he believed that there was an ideal [sic.] ideal self, like Korsgaard does. This is not that important here. My invocation of his views are ment to cast light on the complexity of the psyche and the place of practical identity in all of this.

3 Criticisms

In this chapter I will try to analyse some issues concerning Korsgaard's theory, with the aid of the criticisms by Cohen, Geuss, Nagel, and Williams included in *The Sources of Normativity*. From there I will go to another account that is directed at least in part against Korsgaard, that of Kieran Setiya in *Reasons without Rationalism*, and finally I will air some worries of my own.

3.1 Hobbesian concerns: The issue of authority over oneself

Cohen's criticism of Korsgaard starts out by citing a Hobbesian problem: The Sovereign is above the law, since he can arbitrarily change the laws. According to Cohen, there is a similar problem with Korsgaard's concept of autonomy as a source of obligation; as a self-legislator, you are like a Sovereign, because you can change your mind at any point (in Korsgaard, 1996a, p. 168). Why then, should any 'laws' you make be binding on you? And relatedly, why can't you, as a lawgiver, issue different laws each time to legitimise the action that you are about to perform?

For Korsgaard, in order to change one's mind, one needs a reason. A central premise of Korsgaard's is that we are 'beings who need reasons to live and act', and these reasons, as we have seen, are provided by our practical identities. Our practical identities are reason-providing standards. They come with normative principles, 'laws', that when reflectively endorsed provide us with reasons for action. Once we identify with any given identity, it obliges us to stick with what we happen to identify with. A consistent identity will provide consistent reasons, and without those we lose integrity and identity. Leaving the wanton aside, these reasons can't simply be spur-of-the-moment desires, but reasons connected to desires that can withstand reflection and that we can endorse on a second-order level. As we have seen in the previous section, we must want in general to be the kind of person who does the kind of thing that we do right now, if we are to have Frankfurt freedom.

If the two diverge, we are faced with an obligation to defy our first-order desires in order to stay the kind of person that we want to be in general, i.e. the person we identify ourselves as. Otherwise, we are not in control of our lives, and not really a person at all, at least not a unified agent; in other words, by breaking the laws that we have reflectively endorsed, we exhibit weakness of will, or in more Kantian terms, that we have no will. As for justifying why we ought to be rational at all, looking for the best available reasons, Korsgaard points to the fact that this is a given; it is what we are. We cannot revert to a purely animal way of life, even if we want to. We strive for consistency and integrity, built right into our hallmark ability, reflection. The life of a complete wanton seems impossible, if it is true that our reflective nature forces some minimal reference to reasons on us (Korsgaard, 1996a, p. 99fn). What about leading a 'somewhat' capricious life, would that be possible? Certainly, and indeed it may be the prevalent empirical case, but Korsgaard's contention is that it would leave our identities shattered if it gets out of hand, on account of the lack of integrity and stability of reasons. If we value nothing consistently, then, we lose grip on ourselves as beings who need (consistent) reasons to live and act; in other words, capriciousness is bad for us because we end up valuing nothing, not even ourselves. If we don't have consistent values, we won't know what to do. This would seem to imply the important conclusion that the justification of morality is harmonious with self-interest or well-being, without relying on it.

To answer Cohen's Hobbesian question more directly: laws we give ourselves should be binding because we are necessarily reflective beings who strive for oneness in our lives, which is to say that we cannot help but seek integrity. When guided by reflection and some identity or other, we need what we deem a good reason in order to change our minds. Cohen himself admits that the Sovereign is actually bound by the law until he changes his mind, but he finds changing one's mind to be a trivial affair (in Korsgaard 1996, 170). That we often change our minds in practice, is not something that Korsgaard would deny; what she is getting at is that we should be prepared to back that change up with a better reason that we had before; unless the change is an informed and reflective choice with reference to better reasons, the change will be irrational. Korsgaard's point is not that it impossible to change a mind that is made up – it is rather that there is a best way for one's mind to be made up. And she takes it as a fact that we do seek the best reasons. For our reasons to provide proper integrity they must be based on moral identity, and to arrive there it is required that we push reflection all the way to Korsgaard's source of moral normativity. For now, though, we can stay with non-

moral normativity, and it is sufficient to say that integrity means that we aspire to stick to whatever reasons and standards we have, and those standards are built right into each of the practical identities we happen to subscribe to. And when we break our own laws, which we do from time to time, we will most likely be in pain, in the form of regret or remorse for having acted against our better judgment. Related to regret and remorse, as well as what I just said about capriciousness and the justification of morality, Korsgaard believes that the authority of the mind does not depend on its capacity for self-punishment, but that it 'absolutely implies it. A mind that could not perceive its reasons, after all, could not function as a mind at all.' (Korsgaard, 1996a, p. 151). Pain, in this case moral anguish, is a way to perceive our reasons; and like any other candidate for a reason, it is the subject of endorsement or rejection – though we have a strong tendency to unreflectively reject pain, that is not to say that it is not sometimes warranted.

As for the question of issuing particular edicts to suit the occasion: the reasons provided by those would have to be reasons of an unreflective kind, justified only by immediate desires, and as far as Korsgaard is concerned, this is the weakest kind of justification, as illustrated in 2.5. Cohen offers the example of identifying with a desire to save one's drowning child as a case where it is not required to identify with any law or principle (in Korsgaard, 1996, p. 176). However, if we are queried as to why we did so, we would be pressed to come up with a good reason, however intuitively right the action seems to be. To Kantian eyes, this would be a classical example of our 'moral intuitions' or immediate feelings taking the place of proper reasons, but then we can't really know if the action was right unless moral realism is true and our perception of the situation corresponded to a moral truth 'out there'.

But acting on an immediate feeling in a case like this is not necessarily to say that the action could not have been inspired by the motive of duty – I don't think this motive, or rational considerations, have to be explicit to the agent in the sense that it is 'before his mind' in every action. But on Korsgaard's account we better be able to explain and justify ourselves; when pressed for reasons, we could say that we acted as any good father would do, and with reference to Korsgaard's ethics, one might further say that the laws of the good father is harmonious with those of a good human being, and morally required. When acting on the impulse to save the child, it is a question of identification with some higher-order principle if it is a question of identity at all, some kind of decision about who we think we are. If there is

something like a natural impulse to save the child, that will only make the action easier, but not justified from the point of view of morality.

Like just noted: that the decision should be based on reflective endorsement need not signify the dubious notion that we have to stop and think every time we are about to do something in everyday life. I am not sure whether Korsgaard addresses this, but her view seems to me to be compatible with performing full-fledged actions then and there that are habitually or unreflectively performed, while at the same time being backed by a previous process of deliberation, or self-constitution. In this case, our immediate, strong feeling turns out to be a good candidate for a reason, but it is not always so. Sometimes, when reflecting on actions undertaken, we find in afterthought that by going 'with the gut' we failed ourselves. Perhaps some notion of practical knowledge or practical wisdom is applicable here. With proper experience in doing right, we will know when a suggested course of action requires us to enter a more thorough process of deliberation, while in most cases we might intuitively know what to do on account of previous deliberation, where good reasons can be offered upon request.

Cohen further claims that Kant has an answer to Hobbes' problem of the Sovereign being able to change his laws, whereas Korsgaard does not. The reason for this is that Kant grounds the lawmaking in reason as such, a notion that transcends the merely human, whereas Korsgaard descends to rooting morality in *human* reason, or human reflective consciousness, and that 'all manner of all-to-human peculiarities can gain strength in reflective consciousness' (in Korsgaard 1996, 174). Cohen quotes Kant in *Foundations of the Metaphysics of Morals* as saying that the individual is both subject and author of the law, by virtue of the way he is designed by nature to work as a rational being. In Cohen's own words: 'Kant's person indeed makes the law, but he cannot unmake it, for he is designed by nature to make it as he does, and what he is designed to make has the inherent authority of reason as such' (in Korsgaard 1996, 171). Cohen makes a rather big point out of this not being Korsgaard's argument, since she is 'trafficking at the human level'. His worry seems to be that Korsgaard relies on the notion of practical identity for explanatory purposes. The consequence of all this is that 'Kant can say that you must be moral on pain of irrationality. Korsgaard cannot say that', because for Kant the subject both is and is not the author of the law as well as subject, while in Korsgaard the subject is the sole author of the law (in Korsgaard 1996, 174). As far as I am

concerned, I agree with Korsgaard that if Kant can overcome the Hobbesian problem by referring to the inherent authority of reason, so can Korsgaard. After all, Korsgaard thinks that reflection, the use of your reason, will inevitably lead you to the conclusion that the moral law is authoritative for you as well as all others equally endowed with reason. Her notion of humanity, necessary practical identity and the way it grounds morality is oriented towards the rational part of man, shared by all. Desires are needed to make action possible, but they are to conform to the requirements of reason. Korsgaard is indeed talking about human reason, but after all, can we make sense of any other kind? What is this reason-as-such that Cohen ascribes to Kant, if not the rational part of our humanity, the same part that Korsgaard is concerned about? Could Kant possibly refer to something that is completely beyond us, and would that even be graspable? I think not, but apart from this, I will try and steer clear of Kantian issues to the extent that they obfuscate rather than clarify Korsgaard's argument. I find her argument to do quite well on its own.

3.2 An example of the different levels of reasons and identity

Before we go on, I will try to clarify the delineation in 2.5 by way of example. As Korsgaard introduces the concept of practical identities, she contributes something by way of making us understand our everyday sense of obligation in all sorts of cases, that is, the sense of 'ordinary' obligation that springs from who we think we are. If I find myself in the Royal Castle at a dinner, and my nose is itching, I may analyse the possible nose-scratching morally on account of my viewing myself as a gentleman. Having the identity of a gentleman, of course, will oblige one to follow general rules of etiquette, like not touching oneself above the shoulders, unless with a napkin, at a dinner party and in the presence of others. In this case I would be obligated not to scratch my nose, at least not until I could excuse myself and go to the lavatory. The proposed nose-scratching is then a possible action that I am in a position to endorse or reject. I face the problem of choice: if I am to make a decision, it has to be on account of *something*, some principle.

I might find that I find it fitting according to the categorical imperative, i.e. that I will it to be a universal law that everyone behave as gentlemen to the effect of avoiding nose-scratching at dinner parties. It seems to be the general case that all actions can be normatively potent if they

are analysed in terms of some contingent practical identity. But like we have seen, the categorical imperative does not entail morality in what Korsgaard regards as a proper sense, i.e. one that springs from moral identity and not some contingent practical identity. One might easily imagine some religious identity that prohibits all nose-scratching, even when to oneself, a 'moral' command that would not be sustained if subjected to the standards of moral identity. Korsgaard would call such a command normative, but not morally normative. If, as Korsgaard claims, the categorical imperative does not straightforwardly imply morality, then these are good examples of what one might coherently will as universal laws, but that do not meet the standards of our moral identity or its implication, what Korsgaard calls the 'moral law': 'act only on maxims that all rational beings could agree to act on together in a workable cooperative system' (Korsgaard, 1996a, p. 99) . The point is that any commands may seem justifiable with reference to the categorical imperative alone, if by the categorical imperative we mean only to say the form of a universal law is required.

As Mackie reminds us in his 'argument from relativity', there are a lot of differing moral views out there (Mackie, 1977, p. 36). We create all kinds of different moral laws for ourselves, and it may be that there is no common ground to prove any of them right or wrong. Mackie uses the argument to support his conclusion that moral talk is an error altogether, that it is systematically false. Korsgaard's aim is rather to explain the moral variation by showing that those differing laws stem from contingent identities, but that there is something deeper that can arbitrate between them, recommending some identities, banishing others, or remaining silent where there is no clear impact in terms of the value of humanity.

We could for instance ask whether the nose-scratching at the dinner party has anything to do with Korsgaard's proper morality, that which springs from our moral identity. If I am a Korsgaardian as well as a gentleman, I might find that my identity as a gentleman is at odds with my moral identity. If this is the case, it could mean that the standard of the gentleman should change – like with any practical identity, it is meaningful to discuss what it means to be a gentleman, as the standards involved in different identities or roles seem to change frequently. If the discrepancy between moral identity and the gentleman identity were more severe, they would have to be completely shed on moral grounds. The issue here is complicated because it is one of taste. If I know that the people in the room would find my nose-scratching to be disgusting, does that mean that I am disrespecting them as ends-in-

themselves if I proceed with the scratching? Do they have good reasons to be disgusted? What kinds of reasons? A citizen of the Kingdom of Ends would probably be better off not caring too much about such things, reserving his moral condemnation for more obvious transgressions against others. To me, the philosopher and not the gentleman of the example, it seems that the notion of respect for your fellow man and his autonomy is too thin to provide any clear recommendations in this case. Nose-scratching may be permissible, and as such, it is a question for you to decide whether you wish to stick with that law of the gentleman.

Finally, it is easily imaginable that the scratching of my nose could have clear moral implications, but the context might have to be altered, like scratching one's nose while overtaking a car on an icy road. Universally permitting this would certainly lead to increased loss of life or freedom. One might test a maxim such as 'I will the scratching of my nose in the circumstance of overtaking on an icy road so as to accomplish the relief of my itchy nose', and find that this is not something one would recommend as a universal law, as its general implication is something like 'I will that human loss of life is sometimes acceptable so that people may cater to their negligible desires'. Likewise, we may find that the league of gentlemen – or more plausibly, the religion of non-scratchers – punish nose-scratching in ways that are inconsistent with respecting the humanity of others. If so, they have to go, according to our moral identity. Nose-scratching in seems to fall within the morally permissible, but like everything else, it is really a full description of the particular action as an act-for-the-sake-of-an-end and its motive that matters.

That morality does not have a clear impact in all cases is of course not to say that the concept is empty. In *The Sources of Normativity*, however, Korsgaard only tries to establish the basic value of respect for humanity, and does not delve into its implications. Even so, I may say something about what follows from it; to not deprive someone of their life or freedom seems to be what is most easily and directly derivable from the notion. Incidentally, this corresponds to basic human rights. As for the topic of the morally irrelevant, I will return to it in the next section.

3.3 Cohen's worries about practical identity as normative

For the moment, what is at issue here is the nature of reasons and Cohen's criticism; do all good reasons for action spring from the normative implications of identity? Although Cohen doesn't outline any full-fledged action theory, he definitely thinks not, and emphasises his point thus:

> When I am thirsty, and, at a reflective level, I do not reject my desire to drink, I have, or think that I have, a reason for taking water, but not one that reflects, or commits me to, a (relevantly) normative conception of my identity. Merely acting on reasons carries no such commitments.
>
> (in Korsgaard 1996, 185)

This is offered as a counter-argument to part of his reconstruction of Korsgaard's argument, that 'If we did not have a normative conception of our identities, we could have no reasons for actions' (in Korsgaard, 1996, 185). Recall that we are looking for the best reasons available for action. If we are asking for a *good* reason why Cohen takes his water, I find Korsgaard's views to be able to cast light on what those proper reasons might look like, whether her explanation is right or not. When it comes to Cohen, I am left guessing as to why he thinks he has a good reason. On a common-sense reading, being thirsty is reason enough to drink, but as philosophers, we ask why. Why are you permitted to drink, or: why should you drink? Cohen talks about a reflective level; however, what the nature of his deliberation and the ensuing reason is, he doesn't share with us. Perhaps he is a kind of realist. What seems to be the case is that he has rendered it straightforward that he should drink on account of his thirst, seemingly a first-order desire. It is somewhat similar to the example of the drowning child: it seems that on Korsgaard's account, one can always relevantly ask whether the action is justified. In order to justify your drinking the water, if such a justification is asked for, you will ultimately have to refer to a normative standard, perhaps even a moral one. It does seem a little absurd to demand a justification on account of the trivial nature of the example, but I believe the absurdity stems from it being universally agreed upon that you are indeed morally permitted to drink, unless special circumstances apply – like you being well-hydrated, there being limited water, and your friend dying of thirst – which is not the case in the example. However, it being universally permitted by any reasonable conception of reasons – with or without morality – does not mean that it is not analysable in terms of morality or at least as possibly committing one to a relevantly normative conception of one's identity.

In Korsgaard's world, all actions have to be actions in sense (ii) or (iii) if they deserve the label 'action' at all. So when Cohen states that acting on reasons carries no such commitments, he must be using reasons in another sense than Korsgaard. What Cohen might be saying is that Korsgaard's conception of reasons is far too narrow, not allowing for non-moral reasons or feelings to be proper reasons. His referral to Williams' famous example of the drowning wife (Cohen in Korsgaard, 1996, p. 175) may suggest as much. Williams' view, however, is not something I will deal with yet. At any rate, without a more substantial theory of reasons in place to oppose that of Korsgaard, it is difficult to arbitrate between her view and Cohen's, which leaves me guessing. He needs to state what it is that makes his thirst qualify as a reason, what it is about the situation that has the rightness of taking the drink built into it. He can certainly explain his action by reference to his thirst, but it is not as readily justifiable. Justification makes sense according to a standard, and the standard has not been made clear.

Another suggestion on Cohen's behalf could be that the unpleasantness of being thirsty is itself a reason. The thirst is a mild form of pain, an intrinsically normative fact by some accounts, including mainstream utilitarianism. This common view of pain as a reason is discussed at length by Korsgaard in section 4.3, and the following statement may apply to Cohen's example as well: 'it is not that the pain is an unpleasant sensation that gives the animal a reason to eat. The animal has a reason to eat, which is that it will die if it does not.' (Korsgaard, 1996a, p. 150). Animal issues aside, this statement highlights a difference between a utilitarian view of reasons and Korsgaard's view. I take it as a point to her favour that Korsgaard's theory will be able to accommodate even trivial actions like this, and pass a judgment of reflection in the following way: A desire to drink water presents itself. If we reflect on this, we find that to drink the water is reflectively endorsable on account of the action's end and motive, which would be to respect and uphold the humanity in me, and thus a justified, moral reason. My humanity, like noted in 1.4, rests also on my animal life[24] or health, which is the basis for my continued proper functioning as something higher, a human being, a being endowed with reason. That we don't usually reflect upon whether taking water is a good idea, might only stem from us so self-evidently taking ourselves as important and

[24] For Korsgaard's full argument about your animal nature as an even more fundamental identity that the specifically human rests upon, and the moral status of animals, see section 4.3.10 (Korsgaard, 1996a, pp. 152-3).

worthy of sustainment. Like noted in the previous section, I don't think Korsgaard has to assume that we have to think it trough every time. Having exercised practical reason in a similar situation on a previous occasion should suffice; a law that is laid down stays in effect until you change your mind.

The way I see it, Cohen will have to be clearer about what his positive view of reasons is in order to discredit Korsgaard's view, rather than just stating his disagreement about whether they are connected to practical identity. What seems clear is that Korsgaard shows that they *can* be connected to practical identity, and Cohen fails to convince me that they cannot. That Cohen personally doesn't judge himself to be committed to any practical identity when he drinks water is fine enough, but what he should be doing is offering a reason with more depth than his thirst, or at least an explanation of how that is a reason. By Korsgaard's account, he will be rendered irrational if he fails to offer a good reason for his action.

One further thought; we have seen that the water example can be construed as morally normative. Not all reasons have to be morally normative, though, but it seems they can still be construed to commit you to some normative conception of your identity. Simple preferences like valuing tea seem to fall outside the scope of the apparatus of reflective endorsement and practical identity. But do they? Returning again to the issue of taste and the morally permissible in all this, it doesn't seem like I need to refer to any particular practical identity, at the very least not to my moral identity, to find out whether or not I should take a cup of tea – my likes and dislikes should provide me with reasons enough. Perhaps this is what Cohen should have chosen as an example, not water, as tea (in the general case) has no implications regarding the sustainment of life. Of course, there could be some practical identity denying me the tea, but if real morality has no say, such an identity will be optional at best. Let's say there is no identity to the effect of negatively evaluating tea, rather the opposite – what then? Well, it seems I am free to take my tea, and this seems to apply to all likes and dislikes as long as they don't have moral implications. As it falls outside of the scope of the morally relevant, does it fall outside the model I have made for Korsgaard's idea of practical reasoning? In other words, is there a reason to take the tea that has nothing to do with self-legislation or practical identity, but that is nevertheless a good or acceptable reason? Perhaps something like this is what Cohen is thinking.

To defend Korsgaard's model, we could for instance say that I have adopted the principle of doing whatever pleases me, and that drinking tea is one of those things. Then, I would have the practical identity of an egoist, the 'steward of her own interests' (Korsgaard, 1996a, p. 101). In a sense, we are all such egoists, and some say that is all that we are. Of course, if I am rational according to Korsgaard, my egoism should be constrained by morality, and I would no longer be a proper egoist, but generally speaking a 'moral enjoyer of life' or something like that. So, I drink the tea on account of a contingent practical identity, albeit one that all seem to share, in the sense that everyone has a taste for some thing or other. The implications of the identity are in this case diverse, given that preferences are subjective and different. When it comes to likes and dislikes, they are special in the sense that the egoist identity can contain anything. Maybe it is thus more clarifying to say that each preference is a contingent practical identity if they are not organised under something more general like 'art connoisseur'. There seems to be no contradiction in a 'description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking' construed as directed towards a single preference – after all, if we are to use words like 'tea' at all, they are in a sense general[25]. Being a 'tea-lover' may at least constitute part of why I find my life to be worth living, since it is something I value. If I am a 'tea-lover' then, it doesn't say an awful lot about me and we can assume that it is no part of a core identity, rather way out on the fringe of the total identity (TI), analogous to a non-essential proposition in the coherency theory of truth[26]. Even so, it might challenge my identity in some very slight way to say no to a cup of tea when offered one. I see myself as a tea-lover, you know me as a tea-lover, and I might come to think that 'I am a tea-lover, and that means I should offer you a reason not to take the tea'. For instance that I already had five cups today – in itself an autonomous reason that could be provided by, say, my contingent identity as a person of temperance. I maintain that I am satiated, and nevertheless retain the identity of a tea-lover.

[25] This seems to follow from the universality characteristic of willing, see 2.2.7 below.

[26] Indeed, the coherency theory of truth may be a good analogy in several senses. Where it aims at truth in the form of the integrity of the system, the organization of identities aims at personal integrity and a sense of self. Discrepancy between incompatible identities, like between incompatible propositions, will challenge the system. The consistent failure to act on obligations provided by a core identity can bring down the entire sense of self, like the falsification of a core proposition can bring down the entire set in coherence theory. However, the analogy ends where Korsgaard posits one identity as necessary, dodging the problems of relativism that coherence theory might face. A coherent self in Korsgaard's world, will be modeled around the necessary restrictions provided by moral identity.

The point of all this is that even singular-like preferences can be made to fit into Korsgaard's scheme, and even though this isn't something she worries about when going about her project of creating a foundation for morality, it goes some way to lend credibility to the picture of human agency she presents, even as to include the seemingly morally irrelevant. Particular preferences, some of them sorting under the umbrella of contingent and cultural identities, make up a good part of what we take ourselves to be, and what others take us to be, but intuitively, they shouldn't be able to challenge whether we are a good person if they have no moral impact. If I have odd preferences, it makes me an oddball, but I can still be a good person.

3.4 On Korsgaard failing on her own terms

Korsgaard's theory must succeed in addressing or convincing the radically disaffected; that is the first criterion of justificatory adequacy laid out by Korsgaard herself. Cohen pushes the point that it is too strong a criterion to demand that the alienated person asking the normative question should be convinced by Korsgaard's reply (in Korsgaard 1996, 180-1). I agree, that is an unreasonable demand, if it is conceived as her argument for morality being sure to convince anyone thrown into moral confusion. What Cohen fails to mention, though, is that Korsgaard addresses his problem as soon as she raises the demand, and after a short discussion of the possibility that the other person might not appropriate her answer for various reasons, she concludes: 'for this exercise to work, we have to imagine (…) that this other agent is sincere and reasonable, and does really want to know' (Korsgaard 1996, 16). This meets the criticism by Cohen, but at the same time Cohen seems to be left either not in moral confusion, not reasonable, not really wanting to know, or some combination. Let's assume on Korsgaard's behalf that Cohen is not in moral confusion, the less spiteful of the implications. At any rate, this all points us to an issue briefly raised in the previous section; that a lot of criticism of Korsgaard's theory can be met with an accusation of unreasonableness, irrationality or a blinkered view on a number of issues. Korsgaard thus makes herself immune to much criticism, in this instance by saying that if you are not convinced, the fault lies with you, and not the argument. This is most certainly annoying, but that is not to say that she is wrong. High-handed or not, perhaps hers *is* the best available account of rationality, action, reason and any moral implications that might follow. What we (and Cohen) should ask

ourselves is whether rational demands should be the supreme judge of our actions, and if so, whether Korsgaard succeds in describing what rationality demands of us.

However, there is a lingering element that worries me; in assuming that the enquirer is 'sincere and reasonable', Korsgaard seems to come dangerously close to assuming that the inquirer be *rational*. If this is the case, Korsgaard seems to apply some kind of circular reasoning; if her account of rationality and its implications were true, rational agents would of course be convinced and motivated by it. But if the argument can only convince the perfectly rational inquirer, what is to make us believe that there can be such an agent? If someone were to believe her account of rationality, does that make them rational? No. Unfortunately, belief does not attest to truth. But I believe that this is as close to a validation as we could get; if actual people are convinced by Korsgaard's argument, and take themselves to be acting from a motive of duty and a moral identity, this would at least be some indication of its merits.

3.5 A Mafioso made of straw?

Although Cohen doesn't believe that an 'obligation always takes the form of a reaction against a threat of a loss of identity' (Korsgaard, 1996a, p. 102), he nevertheless provides an example of a Mafioso to argue that even if the connection with identity were the case, this wouldn't have to be the sort of obligation that Korsgaard wants to justify (in Korsgaard, 1996, p. 183). As we have seen, an obligation is a reason not to do something proposed by a desire, upon reflective rejection of that desire. But obligations can be of either type (ii) or (iii) as described in 2.5, depending on whether one applies the standard of a contingent identity or moral identity, respectively.

In Cohen's example, a Mafioso has a 'moral' code of strength and honour, and like the rest of us, he is capable of reflective endorsement. As he is about to murder someone, like he should on account of his identity, he finds that he hesitates. This means that a desire not to kill presents itself - let's call this a 'weak moment' of his. We can then say that he has an obligation (ii) on account of his identity as a Mafioso; knowing that he has to carry the murder through to stay what he is, he reflectively rejects the urge to be nice and sticks with his code. In this way, he retains his autonomy even though he has done something immoral on Korsgaard's account of morality. Korsgaard's answer would be that reflective endorsement

guided by a practical identity is not enough to make an action right – true morality is grounded in our moral identity, with which the Mafioso identity is incompatible.

The Mafioso, then, has used his powers of reflection to steel himself against his immediate impulses, thus retaining his identity as a Mafioso, all the while exercising autonomy. His shortcoming is that he never asked himself something like 'why should my identity as a Mafioso be decisive when faced with a question about what to do?'. Then, he might have come to discover an identity of identities, providing us with the third and final order of reasons, based on what we are that gives rise to identity and value in the first place. Cohen is indeed right that exercising your autonomy is not enough if we are to arrive at morality. But again, this is not something that Korsgaard claims. Korsgaard herself agrees that there is a gap between the categorical imperative (understood as 'choose a law') and the moral law, and makes an effort to fill this gap. Thus, if Korsgaard's argument is successful, the Mafioso poses no threat to *moral* obligation; he just hasn't really pushed his capacity for reflection far enough to arrive at the moral, so his obligation, while there, is of a lesser kind.

Korsgaard claims that autonomy is the source of *normativity*, which is not yet moral. Autonomy, then, is presented as a necessary, but not sufficient component of *moral* normativity. Perhaps there is some confusion here, such as Cohen taking it for granted that Korsgaard is already talking about specifically moral normativity when she talks about normativity. I don't know, because he seems to realise this. Maybe the complaint is that Korsgaard uses the term obligation – even though in sense (ii) – in a way that seems to justify the Mafioso's actions. But his actions being justifiable to *him*, then and there, is not the same as the universal justification Korsgaard is looking for, something that all people should agree on in the full light of reflection. The Mafioso has a first-personal justification for what he does in that he follows his own laws; however, in the light of full reflection, the moral reasons we share as humans would surface, if he were only to arrive there by the use of his reflective capability. The notion that reasons are public is meant to bridge the gap between the first- and third-personal in that our first-personal reasons should be of such a kind that they can be stated and revised in common reasoning. When you fail to arrive at the moral by your own, it is because you don't have access to the best reasons, reasons justified from the common, human point of view, which in Korsgaard's philosophy is as objective as you get; depending on human will, but not subject to whim. Moral reasons are shared by all reflective agents,

even those who have not yet come to realise this, which is why the Mafioso in one sense has his justified reasons – it is as far as he has come – and in one sense not, since he has not yet arrived at the best reasons, the common reasons.

Even so, there could be something more to Cohen's example than what we have seen so far, which at first glance seems to be a complaint that Korsgaard's view has relativistic implications where it does not. What I think Cohen may be getting at is that we intuitively see the Mafioso as rational in his behaviour, without being moral, and that there seems to be no contradiction in this, and thus no necessity of morality following from rationality. This leads us to a problem of a more general kind; is there only one rational law that serves as a basis for all others, and is it moral?

Korsgaard would claim that the reflection of the Mafioso, while operational, is incomplete; in other words, he's rational, but not fully rational. Is this too strong a demand on people, that they should apply reflection to its fullest? I personally don't think that our everyday level of reflection needs to be the guideline when we are looking to ground morality. Korsgaard attempts to formulate what our reasons need to look like if they are to be properly moral, and claims that morality is harmonious with what we are; it is entailed by agency, where immorality is not. However, very few people would fit the description required by Korsgaard for the truly moral person, in that they would not be able to state their reasons like she does. Isn't she in danger of deeming most, if not all of humankind irrational?

Perhaps that is why she appears to be a little vague in her treatment of Cohen's Mafioso; she is not ready to give the Mafioso the harsh criticism that her theory seems to warrant – in fact, she writes that 'I could say that there's no obligation here, only the sense of obligation (…) [but] there is a real sense in which you are bound by a law that you make for yourself until you make another' (Korsgaard, 1996a, p. 257). She further states that it is the endorsement, and not the argument behind it, that does the normative work. Yet, the Mafioso 'ought to have arrived' at the moral, because an essential rule of reflection is that 'we should never stop reflecting until we have arrived at a satisfactory answer' (Korsgaard, 1996a, p. 258). When it comes to the difference between reasons as relate to obligation (ii) and (iii), then, things can easily get muddled, and Korsgaard is only slightly helpful with regard to alleviating the trouble. Are reasons of the second kind moral reasons? Yes and no. They are in some sense

real obligations, and can be taken as moral by the individual. They and based on a certain amount of reflection, deliberation and free choice, but they are still not moral per se. Both kinds of reasons are based on the correct procedure for arriving at autonomy and normativity, i.e. self-legislation by a free will. But the shortcoming of the second kind is that it has not yet reflected its way to proper moral identity as a supreme source of reasons. Humanity as the deep source of values and its implications have not yet been discovered by the reflective mind, and as such, whatever lesser identity prevails may lead to an immoral moral code, like that of a Mafioso. There is no contradiction here in describing the moral code as immoral, given that the code is derived from a contingent practical identity, for instance that of a Mafioso or that of a utilitarian, which can lead to a sense of moral justification, whereas it is immoral under the terms of the necessary moral identity to the extent that the contingent identity in question is in conflict with moral identity. Although Korsgaard is unwilling to say that 'there's no obligation here, only the sense of obligation', what she seems to be saying is that there's no *moral* obligation (iii) here, only an obligation (ii) that may or may not seem moral to the agent, but in any case is a normative obligation of a lesser kind.

At the same time Cohen draws the conclusion that Korsgaard may give the moral person reasons to be moral, while failing to convince the rest, who are nevertheless left with their rationality intact: ' I can show that morality is *a* rational way, without being able to show that it is the (only) rational way', and suggests that this may only amount to a return to Williams' position on morality (in Korsgaard, 1996, p. 181fn). This is a position to which I will return. At any rate, the problem of whether it is possible to be rational without being moral seems to be a question of degree for Korsgaard: the fully rational will question his reasons, even those of contingent practical identities, or more commonly, the reasons of self-interest, which seems to be the standard conception of rationality. But self-interest seems to be able to fit well into the concept of practical identity, because self-interest would have to refer to some list of preferences, or like Williams dubs it, the agent's 'subjective motivational set', that might reasonably be said to follow from what I see myself as. My preferences follow from my self-conception and they matter to *me*, and this gives me a reason to protect them, along with my self-conception. But why should my self-interest have priority? Perhaps an obvious answer is that I value myself above others, which seems to be some kind of fact of life. Korsgaard, on the other hand, would add that while I might tend to value myself more highly than anything else, I have no good reason to do so, because all that I have reason to value in myself is my

ability to value – a property shared by all rational beings. This means that I must value them and their reasons, to the extent that their reasons are proper reasons based on the same insight. All good reasons, i.e. all moral reasons at least, are shared and public. The wanton, the egoist and the practical sceptic are all criticised as ultimately problematic by Korsgaard.

Continuing to explore the question of morality being an option of rationality, we will now turn to Geuss. For one thing, he seems to be of the opinion that reasons can be good, rational reasons without there being any reason for you to adapt them. As we have seen, this may be the case with the morally permissible, but I read Geuss as claiming this may be the case whether or not those reasons are moral (in some sense) or not. He complains about the nonsense in taking others' reasons on as one's own; the 'must' in a sentence such as 'I must value them and their reasons' seems dubious to him. Geuss writes:

> how does it follow from [seeing my humanity as a source of value for me] that *your* humanity must be a source of value for *me*? (…) The Serbs have what I can see are quite good reasons *for them* to act as they do, reasons which (if you will) I can see as arising from their 'mere humanity', but it doesn't follow that these reasons have any standing *for me* (…) [It is] an elementary fact of life (…) that we are constantly encountering people whose reasons for action we understand perfectly well and which we see are genuinely good reasons for them, without in the least endorsing these reasons or sharing their values
>
> (in Korsgaard, 1996, p. 197)

This can be likened to some of Cohen's issues in that it seems to me to amount to a conflation of Korsgaard's two levels of autonomous reflection (ii and iii in my chart). Whether this is on account of a misunderstanding or a flat denial of Korsgaard's arguments for there being a higher level at all is hard to tell. That something is normative from the point of view of some identity is not the same as the claim that it is morally normative. Yes, you can see that there are normative reasons for them to act as they do on account of their identity as Serbs, but at the same time, contrary to what Geuss claims, there is no good reason to endorse them on Korsgaard's account; not from the point of view of morality, insofar as 'Serb' has implications contrary to the demands that spring from Korsgaard's moral identity (iii). Whether immoral or morally irrelevant, their 'genuinely good reasons for them' on account of their 'mere humanity' then, are not-so-good reasons if we refer to Korsgaard's account of the implications of the term 'humanity', in *her* sense. Geuss' sense of 'humanity' here seems to be a reference to biology; at any rate, it is not the same concept as Korsgaard's. The implication of Korsgaard's view is that we must value the others as human beings, as agents capable of reflection, rather than valuing their contingent reasons, which may be bad. The reasons we share come from the moral identity we necessarily share, not the contingent

identities that we don't share. And the rational 'must'[27] in having to value other people's reasons is of course a morally normative requirement of rationality, implying that the reasons in question *should* be likewise based, or at least not in contradiction, if they are to be valued – not a claim that in real, imperfect and irrational human life we 'must' somehow take on any reason the other might happen to have, reflective endorsement notwithstanding. In case the reasons of the other is neither moral nor immoral, e.g. reasons for valuing tea, it seems rationally optional whether we should join in and share those reasons.

What Geuss might be saying here is that moral reasons are either relativistic or nonexistent, if we read him not as attacking Korsgaard's view, but presenting his own. However, what he really does seem to be doing is argue against Korsgaard on what he takes to be her terms, where they are not.

At any rate, this brings us back again to the idea that morality may be no more than an option within morality. If we remove the authority that Korsgaard ascribes to morality, what is left of her view would consist in exactly this: morality is an option within rationality. Rationality would consist in reflective endorsement or rejection of desires on account of them being allowed or disallowed by our identities, morality being a contingent identity equal to the others. Rationality would have to remain silent on what principles/identities to adopt, unless a new theory would legitimise some other organising principle than that of moral identity.

Cohen, while remaining silent on alternative theories, claims that the 'commitments that form my practical identity need not be to things that have the *universality* characteristic of law', assumably to discredit Korsgaard's notion of morality as necessary, or the idea of autonomy. Again, Korsgaard's notion of necessity seems to arouse scepticism. But like I explained, this 'need' for practical identity to have certain characteristics is not a 'must' in the sense that you cannot fail to view it that way. You can easily fail to identify with the general, abstract principle of respect for humanity while still being rationally required to so identify, that is to say that you *should* so identify. As for the universality requirement, I will return to that in 3.7.

[27] Like explained in section 2.2, rational necessity is not a logical necessity, but rather one that produces a 'should' or an 'ought'. Korsgaard writes that the rational necessity of practical reason is the same as that of theoretical reason: 'if I *believe* that all women are mortal, and I *believe* that I am a woman, then I *ought* to conclude that I am mortal' (Korsgaard, 1996a, p. 226fn). You *must* not, in a logical sense, so conclude, but you should if you are rational.

For Cohen as well as Geuss, I think their arguments need to be clearer about both the implication of the terms used and the issues of disagreement with Korsgaard if they are to have a decisive impact.

3.6 Geuss and the implications of undermining morality

Geuss also at one point claims that Korsgaard uses reflection in conflicting senses, one where reflective endorsement or rejection results in reasons and obligations, and one, referring to the discussion of Hume (Korsgaard, 1996a, p. 52) where 'reflection on morality might have a sceptical outcome, undermining the claims of morality' (in Korsgaard, 1996, p. 195). Would a sceptical conclusion imply that we would have an obligation not to be moral? Geuss thinks not, but is unsure about what to do with the observation. I think he is unjustified in his confusion and somewhat wrong that we would have no obligation not to be moral. The latter part depends on the kind of theory that would replace Korsgaard's – in the case of the generic evolutionary theory discussed in 1.1 we could conceivably have a self-interested obligation to deny reasons of morality the priority that they intuitively warrant. Like previously stated, what we have justified reasons for doing, or our obligations, have to be the subject of some standard, so the answer to Geuss' question depends on that standard.

The second sense of reflection that Geuss refers to is theoretical reflection, whereas the first is practical reflection, a distinction that he should be well aware of. The aim of theoretical reason – belief – is a component that should be taken into consideration in practical reason, the aim of which is action. Korsgaard's *Sources of Normativity* is such an exercise of theoretical reason, aimed at convincing us of the truth that morality is real and fundamental, and as such should be a normative force of practical reason, i.e. when deliberating about what to do. What a justified sceptical belief about morality would do, of course, is to abolish the authority of the moral category in practical reasoning, leaving us with contingent identities the contents of which should not be restrained by morality in any straightforward way if we were to be able to reason well. We would then have a (theoretical) obligation – perhaps 'normative reason' is better – not to adopt moral beliefs if we were to aim at the truth. As for action, it depends. Not believing that morality has any special status is not necessarily an obligation to shed moral identity as a source of practical reasons. Like previously stated, it could be an

option within rationality, admittedly a strange one, but it can and has been argued for[28]. Any restrictions on the choice of ends or the means applied in their pursuit would then depend on what further theory of 'best reasons' we come up with. Upon justified moral scepticism, to *believe* that morality was *objectively true* would be nonsense, but what to do with morality as an identity would depend on the adoption of another higher-order organising principle. Morality could for instance be jettisoned as a rational candidate for an identity if justified reasons not to be moral could be found. One could imagine morality as being generally and objectively in violation of the principle of self-interest; or the opposite, that morality was proven as in everyone's best self-interest to adopt as a guideline, truth or no truth. At any rate, Geuss does not seem to have any rationally justified grounds of confusion or complaint in this case, as there is little to be confused about, at least not as far as the two uses of 'reflection' goes.

3.7 Schlegel, Nagel, and the universal character of the will

Geuss cites a Schlegelian critisism of Kant, that true freedom and self-constitution consists in violating one's laws, not sticking with them. The ideal life, according to Schlegel, is 'a constant succession of self-creation and self-destruction' (in Korsgaard, 1996, p. 193). Or, to put it in Nietzchian terms; why not be a Dionysus rather than an Apollo?

The Schlegelian, it seems, would have to be one that doesn't think that there are any best reasons, that they are out of reach, or the like. If this person is acting on reasons at all, it seems they are characterised by a principle of child-like opposition to that which is deemed good. The view of Kant's moral philosophy as the strict voice of reason, entering the scene to spoil fun, enthusiasm and creativity, may be apt here. To the free Schlegelian, it seems the self doesn't need anything but subjective reasons in order to do anything. What then, is left of identity? Who am I now, other than what I am doing right now? If there is any integrity or organising principle to that self, it must be that of opposing the good, or at the very least

[28] See, e.g. Richard Joyce, a contemporary disciple of Mackie in *The Myth of Morality* (Joyce, 2001), on the notion of preserving a kind of moral make-belief after moral beliefs are proven to be systematically false. His position is one of 'moral fictionalism', where morality is kept around not because it is true, but for prudential reasons, i.e. rules of life that it is in one's long-term interest to stick to.

opposing the rational. Korsgaard remarks on the Schlegelian stance of constant self-creation and destruction that in such a condition 'the active will is brought into existence by every moment of reflection, but with no claim to universality, it is no sooner born than dead. And that means that it does not really exist at all.' (Korsgaard, 1996a, p. 232). There is simply nothing left to *be* in terms of identity. It seems that Korsgaard regards the active will as essential to identity, presumably on account of its operation in the reflective, deliberative stance where reasons are established and decisions are made. The reasons we come up with should remain the foundation of action until we relevantly change our mind, which, like we have discussed, should be on account of better reasons. This might be a good time to further elucidate the question of what it means to 'will' in Korsgaard's philosophy, and the relation of willing to the temporal continuity of the self.

Starting from the hypothetical imperative, if you will and end, you must will the means to that end, in the sense that you must be willing means like getting up in the morning and actually be willing to *write* your master's thesis if you *really* will the end of becoming a master of philosophy. As Korsgaard puts it, 'I cannot regard myself as an active self, as *willing* an end, unless *what I will* is to pursue my end in spite of temptation' (Korsgaard, 1996a, p. 231). By acting on the immediate desire to stay in bed when my dog starts barking at six o'clock in the morning, I am not really willing to be a student (or a good dog owner, for that matter). Acting under the idea of freedom, part of Kant's notion of the will as a causality, requires that we must be able to act differently here and now, from the first-personal perspective of practical deliberation. In order to think that you can do something else, you must think about the issue at hand in general terms, that is, you must frame your desires, your means and your ends in general terms to make them comprehensible as options. In conceiving of the issues in general terms, you are at the same time referring to *other possible occasions* than the here and now, occasions where you might end up with the same temptation. All relevant considerations the same then, you will have no reason to act otherwise on another occasion, if you decide that you have a good reason to will such-and-such here and now. This argument is meant to show that in order to properly will, there must be a claim to general validity, i.e. universality, inherent in your willing something[29]. This, in turn, is the same as the categorical imperative

[29] For the most part, this argument is a paraphrase of Korsgaard's in her reply to her critics (Korsgaard, 1996a, pp. 231-2)

being the law of a free will, that is, a will that really wills, and as such rises above determination by the desire of the moment. The claim to universality means law-likeness.

Nagel, echoing Geuss' Hobbesian worries and wondering about why Kant and Korsgaard liken the will to a causality, asks: 'why can't [the will] determine itself in individual, disconnected choices as well as according to some consistent law or system of reasons?' (in Korsgaard, 1996, p. 202). The argument about proper willing means that of course you *can*, but that you would then be a chaos of conflicting desires, and not a self with a coherent identity, not a coherent agent acting on qualified reasons.

What seems to be left of the self on Korsgaard's account is a standing stock of good reasons on account of healthy deliberation, reasons that should be as coherent as possible, and never in violation of the moral law, if the other parts of Korsgaard's argument are right. After what we can imagine as a long process of self-constitution through refinement of reasons, the goal is integrity after the model of Plato's just soul in *The Republic*, i.e. a well-constituted self governed by reason, having a constitution of consistent laws based on the best available and presumably internally consistent reasons. The process of self-constitution, in the dual sense of self-legislation and self-creation, would be a never-ending one, but one can imagine a well-constituted adult to have fewer problems of identity than the adolescent, the adult being past the process of general legislation and dealing mostly with upholding and refining the law. The notion of self-constitution as a final end is the topic of *Self-Constitution: Agency, Identity, and Integrity* (Korsgaard, 2009).

Now, we turn to a philosopher who talks about reasons for action in a more minimal sense than Korsgaard. Kieran Setiya argues that the job of reasons is explanation rather than justification.

3.8 Setiya and the *guise of the good*

In *Reasons Without Rationalism* (Setiya, 2007), Kieran Setiya is out to establish what the title suggests – that we can relevantly talk about reasons for action without what he takes to be the assumptions of ethical rationalism. He also argues that this leaves room for his virtue theory of ethics, not just ethical scepticism. For my purposes of evaluating Korsgaard's theory, I will

focus on the part of his argument that is directed against the *guise of the good* as a doctrine of ethical rationalism.

Setiya begins his book by saying that the 'should' of practical reason is the 'should' of ethics (Setiya, 2007, p. 1). On that much he agrees with Korsgaard. However, in saying that much, he also rejects the instrumentalism of what he refers to a Hobbesian and Humean tradition leading up to modern economics and decision theory, where means-end efficiency becomes the sole criterion of practical reason. He also includes Bernard Williams in the instrumentalist tradition on account of Williams' association between reasons and an agent's subjective motivational set, which I see as correct. He then goes on to reject the Kantian tradition to which Korsgaard belongs because of a notion of practical reason whose 'standards derive from the nature of agency, as such' (Setiya, 2007, p. 2). For the Kantian, the authority of virtues derives from an independent conception of practical reason, but according to Setiya, the standards of reason and the standards of virtue are on an equal footing in the sense that none of them have 'explanatory primacy' (Setiya, 2007, p. 5).

The main force of his attack against ethical rationalism is the claim that it relies on a premise about the nature of intentional action, the doctrine of the *guise of the good* (GG), which he calls a 'normative conception of agency'. The version of GG that Setiya is arguing against is a doctrine about acting for reasons. To act under GG, as conceived by Setiya means that

> acting intentionally is to see some good in doing it;
>
> and to act for a reason is to act on a belief one takes to justify what one does, at least to some extent
>
> (Setiya, 2007, p. 16)

Korsgaard's discussion of action is indeed one of intentional action, as she doesn't seem interested in any other kind. Intentionality must be a requirement of autonomy, since autonomy provides us with reasons and in acting on reasons we are necessarily acting intentionally. In other words, I take Korsgaard to mean that all autonomous action is intentional action. I also take her to be interested primarily in rational action as a way of evaluating intentional action, which may have implications for Setiya's argument. More on this will follow below.

Everyone agrees that in acting for a reason, one is acting intentionally. But Setiya believes that taking something as one's reason has an explanatory, not a justificatory or normative function. He takes this to have falsifying implications for ethical rationalism.

> There is a sense in which, in acting for a reason, one must see the consideration on which one acts *as* one's reason for acting, but one need not see it as a *good* reason for acting, a reason in the normative or justifying sense. (…) one need not see it as doing anything at all to justify what one does.
>
> (Setiya, 2007, p. 17)

In order to reject the guise of the good, Setiya draws on several examples. We can examine two of them, the case of the man estranged from his sexual inclinations and the case of the derisive philosophy student:

> The man who is estranged from his sexual inclinations does not acknowledge even a prima facie reason for sexual activity; that he is sexually inclined towards certain activities is not even *a* consideration' Still, he can certainly *act* on such a desire, and act intentionally – without regarding his reason for acting as any good.
>
> Consider someone who enjoys philosophy for the sense of power it can give, even though he does not see such pleasures as worthwhile in the least. He asks derisive questions at talks because that will humiliate the visiting speaker. This is his reason for acting – he does so intentionally – but recognises all the while that it is not a good reason to act.
>
> (Setiya, 2007, pp. 36-7)

Setiya has several other examples that all share the feature of acting on some inclination which one doesn't seem to endorse, to use Korsgaard's language. Korsgaard, however, favours an account where there is always some minimal sense of endorsement, providing us, it seems, with what I will call a minimal reason, although Korsgaard isn't too clear the impact on reasons of such minimal reflection. Will Korsgaard call this a *good* reason? Certainly not; as already referred to, she explicitly says that reflective endorsement doesn't guarantee any such thing (Korsgaard, 1996a, p. 161). What about from the agent's point of view? Does he have to *think* he has a good reason, on Korsgaard's account? Well, she takes this much for granted: 'when we make a choice, we must regard the object as good' (Korsgaard, 1996a, p. 122). However, this is a basic claim of the relativity of actions to some incentive, and it doesn't imply the stronger claim that we regard our reasons as the *best* available reasons. It only means that the object of any choice, bad or good, must be presented in a favourable light. This may correspond to Setiya's first claim about GG, that 'to act intentionally is to see some good in doing it'. The problem here is that on Setiya's description, the student and the sexually estranged man don't see any good in doing what they are doing. Or do they? Setiya maintains that there is an inclination present. If we understand 'seeing some good' as equivalent to the presence of an inclination, there seems to be no disagreement here. Perhaps this amounts to different readings of the concept of the good, as found in GG; when one sees some good in φ-ing, does that mean that one judges it to be normatively good, i.e. better than the alternative, or does it simply mean that there is some way in which the end of φ-ing is

seen to be desirable? It is the first notion – that one judges it to be better than the alternative – that Setiya is arguing against by way of these examples. Our characters don't judge their actions as desirable, but nevertheless, they must of course desire to do what they do, if they are to act on their bad reasons without being mad. There is something to be said for the presence of a desirability-characterisation here: although the student condemns his actions all things considered, the he still 'enjoys the sense of power [philosophy] can give' (Setiya, 2007, p. 37). The student, then, is rejecting his inclination from the point of view of what he sees himself as having the best reason to do, while nevertheless endorsing it, and acting on it, from some other point of view. This latter point of view can be the point of view of a law that says 'I will do whatever I desire to do', here instantiated as 'I will be detrimental to professors in order to enjoy a sense of power'. The sexually estranged man, however estranged, likewise depends on some desire in order to act on what he sees as no reason at all. As such, we are again dealing with the basic issue of people being sometimes being ruled by whatever desire that presents itself. The student has a minimal reason to be detrimental, and the estranged man has a minimal reason to have sex, even where he sees none. But it is not their best available reasons, and thus don't look like justified reasons, on Korsgaard's account. They could have been justified, in the sense that an action that is affirmative of one's constitution is justified, if they fully endorsed their own actions, but they do not. So whatever the status of Setiya's first claim about GG here, our akratic friends do not *believe* themselves 'to have a justified reason, at least to some extent' (the second claim). Is this a false belief?

If it is true that rationalism presupposes the second part of GG, then akratic or weak-willed behaviour like that of the student and the sexually estranged man would seem hard to explain on a rationalist account. They are definitely, in Korsgaard's terms, breaking the law – their own law. Their actions are akratic because it defies their own evaluative judgment. But do they, to any extent, have a justified reason? If we understand the bad reason the student has, which includes a desire for feeling powerful, as a kind of minimally justifying reason, it might be said that he has a justifying reason, 'to some extent'. Justification, as we have seen, is a question of what standard is applied. In that case, every intentional action is minimally justified on account of its endorsement of some inclination. I believe that to be Korsgaard's position – her view that the action is justified to the extent that it constitutes the agent well and conforms to the hypothetical and categorical imperatives goes together with a view that failure to achieve those requirements amounts to bad action according to the degree to which one fails. So if you understand yourself not to have a justified reason, it might not mean that

your action is altogether unjustified – after all, you see some (lesser) good in doing it, even when you don't think you do – only that you fail to act well. However, things begin to sound a little odd when fitting Korsgaard's views into Setiya's phrasing of GG. This may be because her interest is primarily directed at the question of what *good* reasons look like. In the case where reasons are minimal or bad, one might as well talk about failure of action instead of minimal justification. The justification being as poor as it can be for an intentional action sounds a lot like 'unjustified'. I believe Korsgaard to take something like Setiya's second claim to be true about reasons, but that it is directed at what reasons *should* look like, not what they look like in the worst-case scenario.

Like we have seen, Korsgaard explicitly subscribes to motivational judgment internalism, the view that we are necessarily motivated to some extent by what we judge is the right thing to do. If her position is construed as internalism about normative judgment in general, which I think it must, we are confronted with a familiar version of Setiya's complaint: why we are not doing what we have judged that we should do, when we should (according to the internalist) be thus motivated? But Korsgaard does not subscribe to a *strong* version of motivational internalism, meaning that we are always, necessarily motivated by – and thus, always *act on* – what we judge to be the best reason. It is not clear to me, in the face of the overwhelming examples to the opposite, real-life and imagined, why anyone would hold such a view. At any rate, Korsgaard does not, as it is clear from her formulation of the *internalism requirement* (see 2.3). Rational persons are motivated by their best reasons, not any and every person, all the time. She goes on to say that motivational failure happens in all sorts of cases (Korsgaard, 1996b, pp. 317, 320). So, after all, Korsgaard does not have a problem, at least not *that* problem, with weakness of will, apart from it being irrational, or alternatively, less rational. The student or the sexually estranged man can be said to be rational in the sense that they take the means to some end, only that something has gone awry in the selection of the end. We are irrational a lot of the time, or, like Kant said, we are 'imperfectly rational'. He and Korsgaard both maintain that we have the capacity for rational action, but that this is not something that can be empirically proven.

What are we to make of Setiya's examples, then? Korsgaard's claim will have to be that every agent has the capacity for rationality, and to be motivated by rational considerations to some extent, but that this is far from sure to be the actual case on every occasion. The failure of the two akratics to do what they have decided to do, in terms of rationality, is either a failure of

rationality on their part or a failure of rational considerations to produce a strong enough motivation for action. The demonstration of akrasia as a possibility doesn't quite defeat Korsgaard's theory; after all, hers is a theory about the requirements of rationality, about what we should do insofar as we are rational, which isn't always the case, but rather an *ideal* we should aspire to conform to. The normative strength of one's reason, on such an account, can be evaluated in terms of the rationality one exhibits when reasoning and acting. So while there are cases where it is hard to tell which is the stronger consideration, in real conflicts among the laws one subscribes to, the cases presented here are not such cases. They have minimal reasons, but not good reasons. If when asked, they might say that they have no reason. What Korsgaard takes them to mean is not that they have no reason, but that they evaluate their reasons as bad, i.e. they think they have better reasons not to do what the desire for power or sex suggests. This, however, is just bad action, and not an impossibility on her view. I don't see her claiming that every intentional action has the feature of acting on a good or justified reason, whether so conceived by the agent or not. I only see her claiming that we *should* have justified reasons, as a normative ideal.

Setiya believes the Kantian tradition to assume that 'in acting for reasons, we aim to *justify* what we do', and that this is a mistake (Setiya, 2007, p. 18). I fail to see how Setiya makes the case that the job of a reason is explanation rather than justification. How is the action explained, how is it made comprehensible, if the explanation doesn't involve any sort of justification, however poor? Reasons-explanations are subject to evaluation, to normativity. Setiya claims that '[people] can act for *reasons* we find unintelligible', or that 'one can act for a reason one does not see as good' (Setiya, 2007, pp. 66-7). That may be correct, but only makes those reasons poor reasons, reasons that are justified only in the minimal sense that there has to be some desire present that the agent acts on.

Korsgaard claims that the constitutive aim of action is self-constitution, which includes an aim of justifying our actions. However, this does not mean that it is actually so in all of human practice – Korsgaard believes that if her argument is correct, it shows that we *should* aim to constitute ourselves well. A person, like Schlegel's hero, may possibly not strive for unity, but if that is the case, he fails to be a person with a self at all. What are we to say of Setiya's akratics in light of this? Only that they fail to constitute themselves well, when failing even to stick to their own standards, never mind the standards of rationality. Although Korsgaard would maintain the more trivial claim that when acting for a reason, it is usually

the case that we aim (implicitly, most of the time) to justify what we do and thus constitute ourselves well, that is not the same as saying it is necessarily so, which seems to be the claim that Setiya is arguing against.

Another way of putting it is that Setiya takes the rationalist's *normative* conception of reasons to be a *descriptive* conception of reasons. Where the rationalist, exemplified by Korsgaard, says something about what it is for a reason to be good – it should be maximally, not minimally justified, endorsed in the full light of rationality, or the like – Setiya takes her to say that it is actually what we do when we act – that we are already and always rational agents with reasons we take to be justified to the best of our knowledge.

If I am correct that there is some confusion here, it may be this: Setiya takes Korsgaard to be – like him – talking about intentional action in general, when she is mostly concerned with one particular feature of it, the possibility of rational action. An example is in order to warrant my claim that there is such a misreading. Referring to Korsgaard's *The Normativity of Instrumental Reason* (221), Setiya writes:

> Korsgaard suggests that intentional action is "motivated by […] the *rational necessity* of doing something". To say that we decide upon our reasons is not to say that we decide that they are *good* or *decisive* reasons, just that we decide that they will be the reasons on which we act.
>
> (Setiya, 2007, p. 61)

What Korsgaard is really claiming is that the *rational* agent is so motivated. In a related footnote, Setiya admits that

> the official topic of Korsgaard is "rational action", but I take it that she means by this not "perfectly rational action", but "action of the kind characteristic of rational agents" – namely, *intentional* action
>
> (Setiya, 2007, p. 61fn60)

I think Setiya is wrong to assume that Korsgaard is talking about intentional action in general; to some extent, she is, but primarily, she is offering an account of what rational action should look like. While Korsgaard will have to claim that all rational action is intentional action, I don't believe she subscribes to the inverse claim that all intentional action is rational action, in the sense that it always satisfies the demands she puts on the rational agent. To the extent that the intentional action in question is lacking in rationality, it will look more like any old intentional action the sort of which Setiya is drawing on. However, placing demands on ideal rational action and describing intentional action as it occurs in general are different enterprises altogether. It may not be the case that all ethical rationalists share the same assumptions, but at least in Korsgaard's case, I am reasonably sure about what those assumptions are. I

understand the Korsgaardian guise of the good, like most of her views, to come with a rationality clause; *insofar as we are rational*, taking something as one's reason for action aims at justifying that action, and proper justification at that. Acting against one's better judgment, on bad reasons, or for what the agent himself is prone to dub 'no reason at all', on this view, is not very rational – but it can most certainly be intentional, and like stated, it seems to come with a minimally justifying reason, albeit only in a trivial sense. Korsgaard holds the view that for intentional action, minimal endorsement is always required, and that in this sense all intentional action is autonomous and all autonomous action is intentional, undertaken for a reason even when not meeting all of the standards of rationality. But as Korsgaard says, contrary to what Setiya takes to be her view: endorsement, the 'go-ahead' which is required for action, doesn't guarantee the rightness or goodness of an action (Korsgaard, 1996a, p. 161). When reflection stops too soon, some kind of reason comes out of the implicit endorsement, but it is probably a bad one. Even so, this endorsement implies the responsibility implicit in every intentional action, even when the agents in question don't think it a good idea neither before, nor during, nor after they have acted. After all, they elected to act on that bad reason.

To briefly get back to Setiya's claims about GG and intentionality:

> acting intentionally is to see some good in doing it;
>
> and to act for a reason is to act on a belief one takes to justify what one does, at least to some extent
> (Setiya, 2007, p. 16)

I think Setiya is about right in the wording, but that he is wrong about the implications of his statement because of the mix-up with rational requirements. To see some good in doing an action on Korsgaard's account only implies that a desire has to be present, not that the agent always judges the action done as better than the alternative. And in one sense, every action for a reason is minimally justified by a minimal reason on Korsgaard's account; but again, this doesn't require the presence of a justifying belief understood as a belief about the rightness of the action.

Rational action, however, requires these stronger demands. But as this is a normative ideal, and not a description of intentional action in general, Setiya's criticism misses the mark. Like we have seen, there is a sense in which Korsgaard has to say that we are always acting for reasons when we are acting intentionally; but on her account too, these may very well be bad

reasons. The rationality of such actions is a negative one; the lack of rationality in bad actions still happens within a rational sphere, so to speak, this being the standard that Korsgaard applies to action because of our purported capacity for it.

Setiya seems to fall prey to the confusion of which I've already accused some of Korsgaard's critics, that of mistaking a claim of rational necessity – you must, if you are rational – with the claim that it must actually be so. This is readily understandable; Korsgaard is so focused on rational agency that it is easy to forget that she doesn't claim that all human action is rational beyond the minimal sense. A better line of criticism might be to accuse Korsgaard, with her ideal conception of rational agency, to take us too far afield from what actual action is like. And in a way, this is what Setiya is doing. His starting-point is a broader one: that of intentional action in general, and he presents an alternative account of the relation between reasons, action and virtue that may have its merits. However, I am not about to evaluate those merits; my point here is simply that his argument against the guise of the good is flawed because it ascribes too strong claims to Korsgaard.

Now I will turn to Bernard Williams, a philosopher I believe to have had a great impact on Korsgaard, shaping her approach by way of challenging the Kantian way of thinking about ethics.

3.9 Williams

Bernard Williams' general conclusion about ethics seems to be, like that of Korsgaard, that morality can be justified from the point of view of practical reason, making him some kind of ethical rationalist. At the same time, Williams denies morality any privileged place, making morality an option within rationality rather than a requirement of it. Williams seems to accept rationality as primarily instrumental, thus rejecting the Kantian approach. As we have seen, Williams is an internalist about reasons. Like Korsgaard, he relies on counterfactual claims about what an agent has reason to do, but her notion of practical rationality is not the deciding factor. Williams would say that an action is justified if the agent were to deliberate faultlessly from existing motivations (Finlay & Schroeder, 2008). Unless the relevant motivations are already an active part of our subjective motivational set, we will fail to be morally motivated, and there may not be any way I don't think Williams has any ambition of answering

Korsgaard's morally estranged inquirer, unless of course that inquirer is already someone to whom morality matters.

3.9.1 On the heart's desire

Korsgaard can be accused of alienating us from what we in one sense are; feeling beings whose desires matter. Korsgaard counters this objection to some extent: the identification with your constitution, in view of the soul as a republic ruled by reason where passions have their natural place, is one such effort (see 2.1). Another is the notion of emotion as necessary to perceive one's reasons at all – 'some sort of affect which will direct attention in useful ways is absolutely requisite to getting around in the world at all' (Korsgaard, 1996a, p. 116fn). But at the same time, if you are nothing over and above your desires, you will not be a proper person.

Even so, it seems counter-intuitive to some philosophers that when acting, you are supposed to identify with a principle or your constitution rather than simply identifying with the desire you are acting on. In a lot of situations, it seems apt to identify with or act on the heart's desire alone. Cohen invokes Williams' notion of 'one thought too many' (in Korsgaard 1996a, p. 175). The idea is brought up in connection with a general protest against rationalistic accounts of morality like Korsgaard's, where it looks like reflective endorsement becomes an artificially added element even in clear-cut situations. The example is that of a man being faced with a decision whether to save his wife or a stranger from peril, and

> it might have been hoped by some people (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife
>
> (Williams, 1981, p. 18)

Indeed, on Korsgaard's account, the man might seem to be forced to deliberate like this: he is presented, presumably, with a strong desire to save his wife and a lesser desire to save the stranger. Upon reflection, he finds that it is impossible to save both of them, and that he is morally required to save one. Morality, however, turns out to be indifferent as to whom to save. Korsgaard, with her conception of practical identity, dodges some of the force of Williams' complaint: seeing as the man has other contingent identities to refer to, e.g. commitments pertaining to a husband, it is rationally and morally permissible to choose the wife over the stranger.

But ‘one thought too many’ is having to refer to anything beyond the love for one’s wife in order to justify one’s action. Williams’ point seems to be that theories like Korsgaard’s trivialises what we take to be genuine obligations, perhaps of a non-moral kind, by having to refer to a general principle which seems to distance the agent from the actions he is undertaking. I take Williams to be restating similar concerns, when in his reply to Korsgaard he worries that on her account, something like ‘“I just happen to…” *cannot even intelligently* be applied to’ anything that is supposed to be ‘an adequate normative resource’ (in Korsgaard, 1996, pp. 214-5). Happening to have some contingent identity or happening to love someone are not good enough to provide justified reasons to act on, then. At least not until they have been ‘cleared’ by moral identity. Like I have previously argued, there seems to be some justificatory force in the contingent practical identities, but Williams is right in that they are not completely justified until a moral identity is in place, on Korsgaard’s account. Still, there is a sense in which Korsgaard might say that no reflection is needed, if she maintains the plausible claim that we don’t have to have the relevant reflections before us in our mind at all times (see 3.1, and Korsgaard, 1998, p. 21). The well-constituted agent would feel an urge to rescue his wife because she was his wife, perhaps part motive of duty and part old-fashioned love working together. Analytically tearing apart the elements of rationality and desire when theorising, or when sitting in your armchair deciding who you want to be, doesn’t necessarily alienate us; when something happens, though, we better not just stand there and think.

The element of contingency in all of our identities, however important they are to us, might leave us still wondering why we should endorse them on Korsgaard’s account. One can see the reflective agent deliberating something like this: I need some standard to choose whether to act on an impulse or not. I thus need practical identities, or at least some kind of standards, in order to have reasons. But then, I might continue to ask: Why should I do what my identity, which is also revealed to be contingent, asks of me? After all, I just happen to have a certain family, a certain religion, etc. And then I might reflect my way into depression or at least a very insecure stance on whether there are any justified reasons to do anything at all. Reflection thus proves to be a destructive force if there are no final reasons to be found.

This relates to my earlier invocation of Frankfurt, with his notion of second-order volitions: our will should be determined by the first-order desires that we want to determine it if we are to be free. Frankfurt realises that second-order volitions, corresponding to practical identities,

may conflict, or that we may find ourselves in a position where we have no way of determining what we want our will to be on account on them (Frankfurt, 1997, p. 21). This would render us incapable of finding a desire to determine our will; especially if we come to believe that all is contingent, and there is no firm and final principle from which to act. That is why I proposed that Korsgaard's solution, the moral identity, might provide us with *reasons* of a third order (see 3.5). Relating the quality of reasons to Frankfurt's hierarchy of desires, we could say that first-order desires provides us with minimally justified reasons to act, while the reference to what we want to be, our practical identity, provides us with reasons of the second order and a sense of freedom; that we are what we want to be. Moral identity, on this account, stops further questioning about what we should want to be or what there is good reason to do. It stops questioning by providing third-order reasons, the fully justified kind[30]. It also provides reasons why one should have one practical identity rather than another, but not in all cases, of course. Even the rational agent seems to have options when it comes to the choice between incompatible, but morally permissible practical identities.

Frankfurt, who is not out to prove the authority of morality, suggests that the agent can stop the regress of higher-order desires by 'identifying *decisively*' with one of his first-order desires, declaring that no further questions be asked (Frankfurt, 1997, p. 21). Korsgaard's solution of identifying with your constitution rather than a desire seems to be an attractive alternative that builds on Frankfurt's insights.

3.9.2 On the moral agent's non-moral reasons

Williams airs a further worry related to the previous discussion: 'suppose nothing (relevant) passes the test?' (in Korsgaard, 1996, p. 215). I think by 'relevant' he means our lives, our desires and motivations as they were before we reform them in the light of full reflection, and that there might not be much left of us. Though not brought up here, I take it he refers to problems raised in *Ethics and the Limits of Philosophy*, that of the morally irrelevant being

[30] A curious observation when considering this hierarchy is that the Mafioso in one sense is more autonomous by acting from his identity in spite of benevolent temptation, and thus more justified in his actions than the naturally benevolent type acting without questioning his benevolent desires. Since there is no way of ascertaining whether a person is moral, acting from the motive of duty, or just happens to do the same *act* as the moral type by being naturally virtuous, there can be no separating the two in a legal system, and the naturally virtuous should count himself lucky, but without real merit.

left with a low status, and that one may at the same time find that 'I am under [a moral[31]] obligation not to waste time in doing things that I am under no [moral] obligation to do' (Williams, 1985, p. 182).

What is the status of the morally permissible or irrelevant to be? What would an ideal life look like, is there room for non-moral virtues and joys to be cultivated? What are the limits, if any, of moral obligation, and should we guide our lives and actions by morality alone? A theory that grounds morality seems out of touch with reality if it at the same time banishes life as we know it.

Even so, it is too soon to tell whether Korsgaard's view excludes living one's life in the way we actually do as irrational. She drops us some hints though, without really fitting them into the larger system. During her discussion of Freud and Nietzsche, who were critical of what a moral conscience does to our well-being because of the guilt it creates, she says that 'we can maintain our identities in a general way without each and every moment being ourselves', and that 'maybe a little distance is all we need to keep obligation from getting out of control (…) just as reflective distance gave us control over our animal nature, so maybe reflective distance from our self-control could give us control over it' (Korsgaard, 1996a, p. 160). This is a puzzling thought I would have liked to see her press further. What is it that she suggests here, exactly? That we should allow ourselves minor infringements without feeling overly guilty about it? The notion of reflective distance from our reflective distance screams for an explanation. Maybe it is related to Frankfurt's point that sometimes we need to just make a decision, or we will be paralysed. Or perhaps it hints to the reasonable notion that further reflection about what we are will lead us to conclude that as human beings we are not only rational, but fallible in all sorts of ways, and that complete rationality is empirically unobtainable. So we should forgive ourselves, while striving to be good and to make ourselves better. But attractive as it seems to soften the rational demands, what kind of 'should' is this that seems to qualify the demand for rational action? Is it meant to be some sort of enabling device, to lessen the pressure of obligation and allow ourselves to continue to function as human beings, when the lesser part of that humanity, here extended to include our animal as well as contingent nature, means that we will never be able to fully satisfy the

[31] My interpretation.

demands of rationality? A kind of existential necessity? And when exactly is this supposed to kick in? In the moment of deliberation, which may provide us with an excuse, or just after the fact, the fact being that we have failed and must reflectively overcome our reflective failure? The latter seems more likely. At any rate, the line in the sand between the forgivable and the unforgivable seems increasingly evasive on such an account.

When forced with the task of reinstating normal life as valuable, then, it seems that Korsgaard finds herself way up the creek. Like in the passages just quoted, she is uncharacteristically vague when dealing with Williams' worries in her reply to her critics:

> Must a Kantian regard concerns one just 'happens to have' as 'alien entities from which I must keep my distance'? I agree with Williams that this would be an unattractive result, because we do *stumble*[32] into some of our deepest concerns (…) It is the mark of a kind of immaturity not to accept the deep role of contingency in human life
>
> (Korsgaard, 1996a, p. 241)

She goes on to say that the mature attitude is characterised by embracing contingency, and that this corresponds to Kant's insistence[33] that we should 'take things to be important *because they are important to us*' (Korsgaard, 1996a, p. 242). What is important to us is deeply contingent, and the consequence of biology, psychology and history, and as such, all our values that we happen to hold. By saying this, it seems she is saying the solution to contingent worries is to fully embrace your contingency, like the Norwegian fan of an English football club who has adapted the club as his own in the most contingent way: his father may have been a fan, or the team may have been doing well when he grew up; nevertheless, it has become an integral part of him, enough so that he may conceive it as a kind of necessity. He will experience joy and grief, he will go to the local supporter's pub, and perhaps he will make trips to England to see his team in action. He knows his commitment is random in one sense, but a part of his core identity in another sense, providing him with strong desires and reasons.

Korsgaard thus suggests that the transition from contingency to a kind of necessity be our own work, and like with the football supporter, love between persons can be such a necessity: 'True lovers learn how to be made for each other. Kantian agents transform contingent values into necessary ones by valuing the humanity that is their source' (Korsgaard, 1996a, p. 242).

[32] My emphasis, changed from emphasis on 'do'. She offers the examples of family, nationality, ethniciy, religion, friendship and careers.
[33] A paraphrase with no reference given, just to Kant's 'theory of value'.

So becoming a die-hard football supporter or a true lover is a kind of celebration of humanity, then!

It is still a bit thin, in my opinion. What it echoes is Frankfurt's notion of determining one's will by fiat when one is tired of questioning whether one's priorities are justified. While this has some merits of its own, I am not sure it does much to support Korsgaard's overall system, which is more stringent. Like I noted a few pages back, this appeal to making the contingent necessary seems to be more of a life-necessity than the rational necessity that we have encountered. But like rational necessity, it may in a sense issue from what we are, not just as purely rational valuers of humanity, but as actual, real-life agents that can never reach the fully rational. At any rate, the contingency problem may be overrated. After all, if morality allows it, you are free to do whatever you desire, and hopefully, you desire something. But it can still be a good idea to organise most of what you do into practical identities that you stick with. Lasting decisions provide a firm and healthy self, one might think. And reasons, if we are to believe Korsgaard.

About the question of motivation, Williams remarks that if a moral claim 'can and should have that much power against the heart's desire, it had better have a footing in the heart's desire' (in Korsgaard, 1996, p. 216). In the same vein, he is wondering about how moral identity, upon Korsgaardian reflection, is 'going to make the required elements come alive again' for the person asking the normative question. By the 'required elements' I take it Williams means the motivating desires that are needed to go through with an action, even if the agent in question finds Korsgaard's account reasonable.

3.10 A tale of conversion

Reflecting on Williams thought about how on Earth 'the required elements come alive again', I find myself agreeing with him that it is hard to see how this is to happen. Thus, I want to tell a possible tale of conversion on Korsgaard's behalf.

Seeing as how Korsgaard maintains that the value of humanity is implicit in every choice, one could speak of a process of 'unearthing' your moral identity – because it is already there, implicit in your agency; as far as you choose, you must value; as far as you value, you must value what you are: a value-conferrer; as far as you value yourself, you have no rational

possibility of denying the value of others equally capable – it is as if you say to yourself: 'my God, it turns out I value others by valuing at all', and then discover that you were really thus motivated by acting at all. It is this latter part, that the motivation was really already there, that seems a little dubious. At least if it is cast as a motivation with any decisive force, arising from such abstract considerations. Being motivated from thoughts about your duty may be possible, but it feels like it fails to concern *me*, Korsgaard's earnest, first-personal inquirer. And that is after having read her excellent argument. So how is this to work, exactly?

Even though it is supposed to motivate *me*, I find it hard to use myself as an example, because I am already motivated by moral considerations, and whether or not it is the thought about my duty or just a healthy concern for others from natural or cultural sources, I simply can't tell. Thus, I'll try to imagine that I am someone else.

Let's say I am a typical unlikeable character, the sort of person who hates the fact that other people interfere with my ends. I have selfish but deep-rooted, strong desires to back those ends. How is this artificial-feeling discovery of my moral identity going to change my ways and make me into a gregarious and genuinely moral character? The notion of rationality as a source of inspiration rather than a tool seems at odds with common sense. It has the feel of being taught math; even if you get it, you don't necessarily like it.

If it could happen at all, perhaps it would happen in the following way: I come to accept the explanation as true; I see taking others into consideration as the best thing to do and the moral thing to do. This view of what is best to do will in itself provide me with some motivation. My deep-rooted selfish desires continue to have an impact on me, but I choose consistently to do what I now think is best, what I think is right, in spite of temptation to act selfishly. This will pain me, like going to the gym when I don't feel like it in order to stay with my exercise program; but I do it, because I consider it the better thing to do, all things considered. Time and habituation will perhaps strengthen the motives acted upon and diminish those denied an outlet; even though not strictly necessary as long as I have the motive of duty, these additional motivational factors will make it easier to be good. Acting consistently from what I believe to be the best reasons will strengthen my feelings of self and self-worth. Conviction, then, has reshaped my desires and changed my identity. Having gone through a painful process of change, I have reinvented myself to accommodate my new views about what is right. The related desires have gained force, and those opposing it have receded. I have gained a moral identity, and become, by most people's standards, a better person.

This certainly seems like a *possible* story. But is it probable? And is that important? If we *can* be motivated by the thought of our duty, Korsgaard might have proven her point.

Now, most people love to do what they desire to do, and desire to do what they love to do. They don't go around waiting for some argument or know-it-all to tell them what they *really* have reason to do. Or perhaps they sometimes do, but like we have seen it has to speak to their prior motivations, as Bernard Williams has rightly pointed out, and even then the force of reason to reshape deep-rooted desires and patterns of behaviour seems tenuous. Korsgaard, too, believes that our contingently given ends – issuing from biology, tradition, psychology and so forth – speak to us in an immediate way, but that this doesn't mean that we can't have more complex ways of arriving at ends and reasons that speak to us. The real moral agent will always have a job on his hands if he is to let reason rule, simply because we are so much more than our rational capacity.

This was my personal rendition of what I take to be Korsgaard's story about how moral identity is supposed to make the 'required elements come alive again'. If her inquisitive agent is made too ideal, it is harder to connect with the idea, so I tried to look at it from the point of view of an actual agent who is supposed to be inspired. I believe that in one sense, her arguments in *The Sources of Normativity* and *Self-Constitution* are constructed as a source of motivation for such a process. In another sense, perhaps no such fanciful tale of self-recreation is necessary to validate her philosophical views. After all, these are not self-help books on 'how to become a moral person'. They are intricate philosophical works, probably impenetrable to the general population, intended primarily for the inquisitive and perhaps sceptical academic who wonders how a foundation for morality can be argued for in our Godless age. The academic may be moral in a conventional sense, but question why he should be. And so we get Korsgaard's argument to that effect, presented as addressing the genuine but disaffected inquirer. Is he not us, the academic community, rather than my unlikable fellow, who most likely will not reform himself given any theoretical resource? What I see the argument as doing, is to make an honest attempt at answering the normative question; it is an attempt to verify that we are, insofar as we have moral convictions, justified to have them. And I believe what I take to be an assumption of Korsgaard's, that we seek to be as justified in our actions as possible most of the time. But it could still be the case that ultimate justification is an impossibility.

We *should* have a moral identity, Korsgaard argues, because it is the rational thing; it is in a sense the solution to the problem of choice. What follows from the argument, if successful, is a justification of action towards the construction and preservation of moral institutions, human rights and moral education. Not an argument guaranteed to persuade and change any old amoralist, who may very well be beyond rescue – one might argue that his moral identity is there only in the counterfactual sense, not to be awoken, even if there is a possibility of some or most humans having that possibility from birth. In the face of a good and convincing argument, the real agent might not be convinced for so many reasons, and even if convinced, it is an odious task, to change what one is. We can only do so much after life's contingencies have had their way with our personality, which is why moral arguments should aim at convincing those already somewhat convinced to redouble their efforts to make society better.

All in all, I find myself siding with Mill on this one (see 1.2). Not on the part about utilitarianism, but on the part about the power of arguments. Korsgaard seems to be right that arguments sometimes convince and motivate, but on their own, I don't think they are enough. We are simply not rational enough, even if Korsgaard is right about rationality as an ideal and its moral implications. Rational considerations about agency may possibly be able to motivate, but what is dead sure is that all our contingent desires are so able. So in light of being convinced by an argument like Korsgaard's, and finding a moral life to be a good life, we can perhaps still do what Mill suggested and bring up our children to become moral, making as much out of nurture as possible. We can secure social institutions that further the value of human life and the human capacity for rationality as well as the capacity for more silly, but harmless things. We can provide an example for other cultures; we may even put pressure on them, while still arguing with them and believing in the power of argument to motivate to some extent. Even the bad father, if convinced by the argument or otherwise motivated, will not want his daughter to become what he is, though he may not be able to change himself even when convinced about his shortcomings.

4 Conclusion

Rational necessity has proven to be a source of confusion that has lead some critics astray. But even when there is less confusion, it is not readily evident that we are moral on account of our agency or that we are capable of being motivated by such considerations.

Especially the part about having to value yourself and then having to value others seems questionable. The argument is interesting enough, perhaps even strong enough, but the motivational source seems very abstract. Its abstractness is accompanied by a disheartening realisation: whether we ever be act from rational considerations or from ulterior motives cannot be verified:

> The extent to which people are actually moved by rational considerations, either in their conduct or in their credence, is beyond the purview of philosophy. Philosophy can at most tell us what it would be like to be rational.
>
> (Korsgaard, 1996b, p. 332)

At the very least Korsgaard has made a powerful attempt at the latter; telling us what it would be like to be rational. She has provided us with an extremely rich argument both about our contingent sources of normativity and about our purported necessary nature that may serve as their regulating ideal, creating a foundation for morality. I have tried to make that argument as clear as possible by exploring it from different points of view and questioning it from some of them. What we are left with is inconclusive, but enriching.

Bibliography

Finlay, S. & Schroeder, M., 2008. *Reasons for Action: Internal vs. External.* [Online]
Available at: http://plato.stanford.edu/archives/fall2008/entries/reasons-internal-external/
Frankfurt, H., 1997. *The Importance of What We Care About.* Cambridge: Cambridge University Press.
Joyce, R., 2001. *The Myth of Morality.* Cambridge: Cambridge University Press.
Kant, I., 1785, 1903. *Grundlegung zur Metaphysik der Sitten in Kants gesammelte Schriften.* Königlich Preußische Akademie der Wissenschaften ed. Berlin: Reimer.
Korsgaard, C. M., 1985. Kant's formula of universal law. *Pacific Philosophical Quarterly 66, no. 1-2,* pp. 24-47.
Korsgaard, C. M., 1996a. *The Sources of Normativity.* Cambridge: Cambridge University Press.
Korsgaard, C. M., 1996b. *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press.
Korsgaard, C. M., 1998. Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind. *Ethics,* 109(1), pp. 49-66.
Korsgaard, C. M., 2003. *Normativity, Necessity and the Synthetic a priori: A response to Derek Parfit.* [Online]
Available at: http://www.people.fas.harvard.edu/~korsgaar/Korsgaard.on.Parfit.pdf
Korsgaard, C. M., 2009. *Self-Constitution.* Oxford: Oxford University Press.
Mackie, J., 1977. *Ethics: Inventing Right and Wrong.* s.l.:Pelican Books.
Nagel, T., 1970. *The Possibility of Altruism.* Princeton: Princeton University Press.
Rogers, C., 1961. *On Becoming a Person.* Boston: Houghton Mifflin.
Setiya, K., 2007. *Reasons Without Rationalism.* Princeton: Princeton University Press.
Williams, B., 1981. *Moral Luck.* Cambridge: Cambridge University Press.
Williams, B., 1985. *Ethics and the Limits of Philosophy.* Cambridge, MA: Harvard University Press.
Wittgenstein, L., 1953. *Philosophical Investigations.* Oxford: Blackwell.