

August 2013

Commitment and Temporal Mediation in Korsgaard's Self-Constitution

David Shope

University of Wisconsin-Milwaukee

Follow this and additional works at: <http://dc.uwm.edu/etd>

 Part of the [Ethics and Political Philosophy Commons](#)

Recommended Citation

Shope, David, "Commitment and Temporal Mediation in Korsgaard's Self-Constitution" (2013). *Theses and Dissertations*. Paper 224.

This Thesis is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact kristinw@uwm.edu.

COMMITMENT AND TEMPORAL MEDIATION IN KORSGAARD'S *SELF-
CONSTITUTION*

by

David Shope

A Thesis Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Master of the Arts

in Philosophy

at

University of Wisconsin-Milwaukee

August 2013

ABSTRACT
COMMITMENT AND TEMPORAL MEDIATION IN KORSGAARD'S *SELF-
CONSTITUTION*

by

David Shope

The University of Wisconsin-Milwaukee, 2013
Under the Supervision of Professor Edward Hinchman

In *Self-Constitution* Christine Korsgaard argues that our reasons are public. What she means by this is that if a rational agent has a reason to perform some action, it is a reason that has normative force for everyone who is a rational agent. Korsgaard also argues in *Self-Constitution* that when we will a course of action, we must do so in the form of a determinate commitment. Doing so requires determining some reasons to be bad reasons to opt out of the course of action that we will. Finally, Korsgaard claims that the selves occupying our own body at different times are distinct agents unless their wills are unified. In this paper, I will argue that Korsgaard's views about diachronic identity produce tensions between her claims that reasons are public and that volition involves determinate commitment. If reasons are public, then my future self's reasons whatever they may be cannot be preemptively dismissed as bad reasons. Yet, in order to commit ourselves to a determinate course of action through our wills, Korsgaard claims that this is precisely what we must do. The only way for Korsgaard to resolve this conflict between her claims is to argue that the form of commitment she describes is a necessary

form of mediation between the reasons of agents occupying the same body at different times. I will consider an argument that mediating in this manner is necessary for the efficacious pursuit of our ends, and therefore required by the constitutive features of agency. I will show that this argument is unsuccessful in establishing that such a strategy of diachronic coordination is required to pursue our ends and that, further, such a strategy will impinge upon autonomy of agents subject to it since it allows the deliberating self to arbitrarily establish restrictions on the reasons its future self might be motivated by.

TABLE OF CONTENTS

Introduction.....	1
I. Agency as Self-Constitution	5
II. The Publicity of Reasons: An Overview	7
III. Diachronic Cooperation and the Categorical Imperative.....	11
IV. The Argument from Deliberative Neutrality.....	18
V. Pluralities of Agents and Pluralities of Principles.....	21
VI. The Final Step	25
VII. Anticipating Reasons.....	34
Conclusion	49
Bibliography	50

Introduction

Christine Korsgaard has developed a Kantian account of agency that places sociality at the foundation of all rational action. Her account of diachronic action most vividly portrays this: on Korsgaard's account the self that wills some action at a time and the self that carries it out at another time are not the same self, even if they share the same body, except inasmuch as their wills are unified. The selves occupying my body have to share projects if they are going to get anything done. This is not for Korsgaard just a thesis about the persistence of personal identity. It plays a crucial role in her argument that the Categorical Imperative is constitutive of action.

More broadly, Korsgaard takes reasons to be *universal* or, in her terms, *public*. Any reasons aren't just mine, they're yours too and vice-versa. This comes out most strongly when Korsgaard compares the lack of privacy of reasons with Wittgenstein's argument against private language. Korsgaard suggests that normativity permeates the social to such a degree that "by calling out your name, I have obligated you, I have given you a reason to stop."¹ In my life as an agent, I encounter reasons all around me, reasons that are reasons because *others* take them to be. It is not merely that reasons are public in the same way the world is public, there to be found by anyone who would look. Rather, Korsgaard thinks that reasons result from our reflective endorsement of maxims for action and so reasons are only to be found in the domain of interacting rational agents. I will describe her account in somewhat more detail in Section I and then most fully in Section V, but it is important to bear in mind that Korsgaard thinks normativity is literally willed into being by rational agents, albeit with certain constraints on what counts as volition.

¹ Korsgaard 1996, 140.

In the following sections I will attempt to clarify Korsgaard's argument for her claim that reasons are public in this manner. Korsgaard argues in three steps. The first two steps of her argument for the publicity of reasons establish a certain kind of public availability of reasons but do not support her stronger claims that reasons are public in that they make demands of us by default. The first part of the argument establishes the claim that *my* reasons must be formulated in a manner such that they are available to others by taking on the law-like form of a determinate commitment, thus grounding the possibility of volitional unity. The second part establishes the claim that there is no basis on which the incentives on which others with whom I am interacting are operating can be excluded as potential grounds for my own practical deliberations and that, therefore, there is no pre-deliberative class of reasons which is uniquely mine. Both of these steps ensure that public reasons are *possible*. In the final step, Korsgaard argues that it is necessary for our reasons to be public since that is the only way to interact and interacting is necessary for any action at all. Here, Korsgaard claims that in order to cooperate with other agents, we must deliberate from a standpoint of mutual respect for our capacity to legislate (what Korsgaard calls our humanity). She then suggests that because, as she shows in the first step of her argument, all action requires cooperation with at least future selves occupying one's body we must always act from a respect for humanity and that there is no way of privately demarcating humanity to respect: it belongs to all agents. Thus, since respect for humanity requires treating others' reasons as normative for us, we must reason publicly if we are to act at all. The result is that, much in the same way that through friendship or marriage each participant takes on the other's projects as their own, we must take on the

projects of other human agents as our own because we are constitutively committed to treat our wills as unified with theirs through our common humanity.

However, a concern arises once Korsgaard has got her argument on the table. She appeals specifically to the case of diachronic action in order to argue that reasons are public because diachronic action is inescapable. Her model of public reasoning as shared deliberation seems like it could not apply in the diachronic case: past and future selves are not present to each other in a way that would allow them to deliberate together and come to a shared conclusion which they endorse simultaneously. Therefore, some sort of policy needs to be in effect governing how, across time, we ensure that we act in a coherent manner while still respecting the publicity of reasons. In the context of temporal separation there appear to be only two options: either the past self lets the future self decide, which undermines the determinacy of commitment which Korsgaard presents as an essential feature of volition in the first step of her argument or the future self lets the past self decide, which seems to require that the determinate commitments we create through our volitions can justifiably preempt the judgments of one's future selves. Furthermore, whatever policy binds us across time it must be either one or the other. The policy in which future self lets the past self decide cannot gain its authority from any past self's commitment to it: otherwise, one would antecedently commit the future self to accepting the authority of antecedent commitment. Thus, if our volitions are determinate commitments they must be so constitutively: it must be a part of the nature of volition that it takes the form of determinate commitment.

I will argue that our volitions being determinate commitments cannot, on Korsgaard's account, be constitutive of our agency. There is some appeal to the policy of

treating our volitions as committing us inasmuch as doing so would give us normative authority to pursue our ends in a temporally extended manner even in spite of preference shifts which might constitute competing judgments. However, this commitment is not strictly necessary in order to be able to act. This is most plain, I will argue, when we consider that we are capable of acting across time even when unanticipated circumstances place us outside the realm of circumstance in which our commitments tell us what to do. Since we are finite agents, our capacity to commit ourselves is always limited. If commitment is necessary for action, then we are frequently failing to act.

The outline of this paper is this: in Section I, I will give an overview of the foundation of Korsgaard's account of agency; In Section II, I will present Korsgaard's account of public reasons briefly, so that the progression of her argument is clearer. In Section III, I will present the first step of her argument; In Section IV, I will present the second step of her argument; In Section V, I will argue a crucial point: that for Korsgaard, we must unify not only our incentives under principles of volition, but our principles of volition into a coherent whole. The result is that the unification of wills is an all-or-nothing affair, even in cases of interpersonal interaction. In Section VI, I will present the final step of Korsgaard's argument. Lastly, in Section VII I will consider whether or not volition is constitutively determinate commitment and argue that it is not.

I. Agency as Self-Constitution

I will begin with Korsgaard's account of non-human animal agency, as it helps illustrate the way in which Korsgaard thinks our agency is tied up with principles of action. On Korsgaard's account, animals represent the world through perception in a way that construes the features of that world as calling for certain responses: "The world as perceived by the animal is organized around his interests: it consists of the animal's food, his enemies, his potential mates, and, if he is social, of his fellows, his family, flock, tribe or what have you."² Animals possess instincts which manifest as dispositions which lead it to perceive and respond to the environment in a certain manner, governed by the function of that animal: the preservation and reproduction of its form.

However, in the case of human beings there is another level at play in the relationship between our representations of the world and our actions. According to Korsgaard, we human beings "are aware not only *that* we desire or fear certain things, but also that we are inclined to act in certain ways on the *basis* of those desires or fears."³ This results in what Korsgaard calls reflective distance, the space in which questions of justification arise. We have to decide whether or not we count the incentives presented to us by our instincts as reasons. The result is that "instincts no longer *determine* how we respond to those incentives, what we do in the face of them."⁴ Since the instincts no longer determine our actions, Korsgaard claims we need principles, what she calls maxims, in order to settle what will count as reasons. These principles replace our instincts in constituting our form as agents.

² Korsgaard 2009, 110

³ Ibid, p. 116

⁴ Ibid.

Korsgaard argues that there are two aspects to human agency under this description. First, the principles of action we endorse must take the form of a law. To will a principle as a law means that when we will that principle we are making ourselves into a cause which will bring about the goal contained in our principle.⁵ In order to do this coherently, we must will the means to that goal as a part of the action we undertake. Korsgaard takes these features of actions to be expressions of the fact that the categorical imperative in the formulation of universal law is constitutive of human action. In other words, we have to act on principles possessing these features in order to act because on Korsgaard's account, those very principles form our make-up as agents. This is the fundamental constitutive norm of human agency, but it is a merely formal constraint. We must also have particular principles that we adopt directing us at particular ends. This is the second feature of our agency, tied to our animal nature. The positive content of our actions must come from outside the formal constraints of our agency. In fact, Korsgaard claims all our *non-moral* values come from our incentives, which is to say our instincts. Thus, as agents we are still working with the materials that our constitutions as animals provide.

The reason why Korsgaard thinks we need principles is that when the grounds for action naturally provided for us by instinct fall away in the face of reflective distance, we need new grounds for action. Since our actions are no longer immediately governed by our experience of the world as organized around our preservation as animals, we have to reconstitute ourselves as causes under a different form. However, in the case of human

⁵ The reason 'being-the-cause-of' is connected with lawfulness on Korsgaard's picture is its Kantian heritage. In the same way that natural laws define causal relations, the laws we give ourselves are supposed to define causal relations. When we give ourselves laws *we* are the cause.

beings Korsgaard thinks our forms are up to us. Therefore, rather than operating on instinctive principles of action, we formulate maxims which represent what courses of action we reflectively endorse. The principles of action we endorse are what constitute us as agents. The reason why we must endorse *principles* rather than, say, one-off behaviors is that Korsgaard thinks there is no other way to constitute one's self as an agent at a time unless one endorses a principle which is available to all rational agents. It is in her argument for this claim, the claim that we must act under laws we give ourselves that the argument for public reasons begins.

II. The Publicity of Reasons: An Overview

In this section I will provide an overview of Korsgaard's claims that reasons are public, primarily by considering her discussion of the issue in the chapter "Integrity and Interaction" from *Self-Constitution*.

In the chapter "Integrity and Interaction" Korsgaard provides an account of public reasons, arguing that in interactions with other agents reasons are shared. In the context of a violent property dispute, Korsgaard illustrates the way the universalization requirement of the Categorical Imperative differs between private and public accounts of reasons:

I think I have a reason to shoot you, so that I can get the object. On the private conception of reasons, universalizability commits me to thinking you also have a reason to shoot me, so that you can get the object... but on the public conception of reasons, we do not get this result. On the public conception I must take your reasons for my own.⁶

Naturally, willing both that I shoot you and that you shoot me in order for me to get the object produces a practical contradiction and so I cannot will to shoot the other person in order to obtain the object on a public account of reasons. Korsgaard's claim here boils

⁶ Ibid, 193

down to the following: "if personal interaction is to be possible, we must reason together, and this means that I must treat your reasons, as I will put it, *as reasons*, that is, as considerations that have normative force for *me* as well as you."⁷

Korsgaard makes stronger assertions later: "We can't choose to treat someone's reasons as reasons, as considerations with normative force for us... responding to another's reasons as normative is the default position--just like hearing another's words as meaningful is the default position."⁸ It is this stronger articulation of the publicity of reasons that will be the main issue in this paper.

One important feature of Korsgaard's account is that interaction occurs in all cases of diachronic action. At minimum, all action across time is a kind of interaction among successive agents occupying a body. Korsgaard puts the matter quite plainly:

The requirements of unifying your agency internally are the same as the requirements for unifying your agency with that of others. that's why you have to will universally, because the reason you act on now, the law you make for yourself *now*, must be one you can will to act on again *later*, come what may, unless you come to see there's a good way to change it.⁹

For Korsgaard, diachronic action of any sort requires public reasoning. For, one is not a united self until after the will is united across the selves occupying one's body and wills can only be united through the recognition of public reasons. To illustrate, Korsgaard discusses the scenario of the Russian nobleman.

The Russian nobleman, in his youth, is confronted with a difficulty: while at present his ideals and values suggest to him that he should redistribute his wealth to others, he knows that as he becomes older he will become more conservative. He

⁷ Ibid, 192

⁸ Ibid, 202

⁹ Ibid, 203

undertakes to establish a contract, which he leaves in the control of his wife that will bind his future self to undertaking the plan of redistribution. On Korsgaard's account, this sort of behavior can only express a breakdown in volition: the Russian nobleman "expects to change his mind without a reason."¹⁰ Hence, the nobleman cannot be a unified agent in the undertaking of the wealth redistribution. There is a failure of action present.

Korsgaard describes this failure in different ways: either as the failure of the young nobleman to take his older self's reasons as his own or as a failure of the older nobleman to take the young nobleman's reasons as his own. Both are, assuming the young nobleman's prediction is correct, failing to recognize reasons as public. Thus, for Korsgaard the very structure of diachronic action requires the publicity of reasons because interaction requires the publicity of reasons and all diachronic action just is a kind of interaction.

It is important to note that Korsgaard takes the claim others' reasons have on us to be defeasible: "[The young nobleman] can decide to disagree with his own future attitude. But *unless* he is then prepared to regard his own future attitude as one of weakness or irrationality, he is not according the reason he himself proposes to act on *right now* as having a normative standing."¹¹ If there is reason to doubt another agent as a rational source with respect to reasons, then I need not take their reasons for my own. This opens up the question as to what amounts to being mistaken or irrational on Korsgaard's account. This means that to understand the sense in which Korsgaard wants there to be a default inclusion of other agents' reasons in my own deliberative process, we will need to settle her account of the norms of agency. Perhaps unsurprisingly, Korsgaard

¹⁰ Ibid

¹¹ Ibid

takes the fundamental norms of agency to play an important role in grounding the publicity of reasons, so to even have a full grasp of what Korsgaard means by calling our reasons public one must understand this feature of her account. The publicity of reasons is an expression of their universality, which is a result of the fact that agency not only must comply with the categorical imperative¹² but is constituted by reflective endorsement constrained by it. Korsgaard not only justifies but characterizes the categorical imperative in terms of the same sociality that underlies not only cooperation with others, but cooperation with the selves occupying one's body across time. The link is with efficacy in action, which amounts to efficacy in the formation of unified volitions.

Here we can see the issue that will concern us later in Section VII: exactly how do our diachronic self-relations work? What are the defeasibility conditions exactly? For instance, what does it amount to for the Russian nobleman to consider his future attitude as one of weakness? What role does the Russian nobleman's anticipation of his older self's reasons play in his deliberation? On the one hand, Korsgaard seems to suggest that the Russian nobleman must cooperate with his future self unless his future self's judgment can be seen as irrational. But the only norms governing the rationality of agency on Korsgaard's accounts are the norms governing self-constitution. Thus it seems that the Russian nobleman cannot commit himself to giving his wealth away unless he thinks that to do otherwise would be in violation of the constitutive norms of agency. Yet, as we will see in the next section, along the way to arguing for the publicity of reasons, Korsgaard suggests that commitment is constitutive of volition.

¹² Korsgaard thinks that the hypothetical imperative is not a distinct principle from the categorical imperative, so inasmuch as the features of the categorical imperative are most relevant for her account of the publicity of reasons, I will only be focusing on it.

III. Diachronic Cooperation and the Categorical Imperative

To be an agent for Korsgaard is to be autonomous: to give one's self laws. On Korsgaard's account the actions we will contain an end goal and the required means to get at that end goal. An action takes the form of what Kant would call a maxim: "to-do-x-for-the-sake-of-y." This is the general form of any willed maxim. Maxims are principles of action, and we need maxims because we are reflective: we need to give reasons and justify when deciding on what to do. The reason we need reasons, according to Korsgaard, is because our self-consciousness allows us to reflect. Without self-consciousness, we would be moved to behave as a result of any incentives presented to us but "it is within the space of reflective distance that the question whether our incentives give us reasons arises."¹³ This question arises because the reflective standpoint is one of justification. Now, in order to provide justification "we need principles, which determine what we are to count as reasons."¹⁴ Without reflectivity, we don't have justification and so normativity isn't even in the picture yet. But, for Korsgaard we need principles in order to answer the challenge of justification. Those principles are maxims.

For Korsgaard there are two key features of maxims. First, maxims take the form of a law. Second, they are universal. To will a maxim as a law means that when we will that maxim we are making ourselves into a cause which will bring about the goal contained in our maxim. In order to do this coherently, we will also have to will the means to that goal as a part of the action we undertake. The maxim that we will in undertaking an action must also be universal, which is to say that it should be able to be willed by any and all rational agents. Korsgaard takes both these features of actions to be

¹³ Korsgaard 2009, 116

¹⁴ Ibid

expressions of the fact that both the hypothetical and categorical imperatives as Kant described them are constitutive of action. The claim that the categorical imperative is constitutive of action is crucial in support Korsgaard's ultimate argument that reasons are public. If actions were not governed by principles, as the categorical imperative demands, then reasons would not be able to be shared in the first place since there would be no possibility of multiple agents acting on the same maxim. The generality of the principles of volition which the categorical imperative demands is precisely what allows for multiple agents to adopt the same principle and therefore to share their reasons.

Korsgaard, therefore, argues that our maxims must take the form of a universal law, which is to say that they must conform to the categorical imperative. The maxim Korsgaard takes as an example is that of going to the dentist to have a cavity filled. Willing the maxim universally means that "I commit myself to acting as this maxim specifies -- going to the dentist on the occasion of my appointment --... so long as I still have both the cavity and the appointment, and unless there is a good reason why not."¹⁵ In order to be universal the maxim has to somehow generalize.¹⁶ In spite of changing conditions, I commit myself to acting as the maxim demands assuming the relevant requirements are met. In this case, one of the relevant requirements is having a cavity. However, Korsgaard thinks there can be others. This is her motivation for including the italicized clause. The clause "unless there is a good reason why not" expresses that maxims we will are what Korsgaard calls 'provisionally universal.'

¹⁵ Korsgaard 2006, 61

¹⁶ Korsgaard uses the term 'general' to describe a particular sort of principle which is distinct from universal principles. This technical use of the term will be irrelevant to our discussion, so my use of it will be that of ordinary English.

If it is the case that willing maxims as universal laws is constitutive of action as Korsgaard claims then in order to will to go to the dentist to fill a cavity I will have to will my maxim by committing myself to it as a provisionally universal law. But why should I do that? According to Korsgaard, it is because if I don't will it universally I can't be said to have made a commitment and if I haven't made any sort of commitment to my maxim "then I have not really willed anything."¹⁷ For instance "it may be that I am really terrified of the dentist and therefore I am *always* tempted to find some excuse not to go when the day arrives. Now if I am prepared to give up the project of going to the dentist in the face of *any consideration whatever* that tempts me to do so, then clearly I have not really committed myself to anything."¹⁸ My maxim has to generalize across potential situations such that it applies to some situations and not others. It has to have determinate contours that establish when it is appropriate to act in certain ways and when it is not. Otherwise, it would make no determinate demands and would not amount to a commitment of any sort. Once my maxim has this universal character it becomes available to other agents. It is a rule they could potentially follow just as much as I could since its generality allows it to apply just as well to situations they may find themselves in. Put another way, it is not limited to just *my* particular situation because my particular situation isn't going to persist. In virtue of my maxim's generality it is, in a sense, public inasmuch as others can follow the same maxim. Korsgaard also argues that unless my maxim or reason can be *available* to any other rational agent as a potential normative standard in this manner, there can be no cases of volitional unity across time. The reason

¹⁷ Korsgaard 2006, 61

¹⁸ Ibid, 62

for this is that Korsgaard maintains that "the self is constituted by volition."¹⁹ So, "when I will to go to the dentist on the day of my appointment, I cannot be willing a law that *my future self* should go to the dentist, for whether I *have* a future self depends on whether that law and others like it are obeyed."²⁰

This sudden talk of future selves might seem out of place, but its appearance here is quite natural. Korsgaard can best demonstrate that our maxims must be available to other agents by demonstrating that even when our volitions do not stray any further than home-base (our body) they are required to be available to other agents. Otherwise, a counterexample to her contention that all maxims are available in this manner would be cases where I will a maxim such that I only need to directly involve *myself* in following it and it seems that many candidates would be available if I had an independently persisting self. If all cases of action require maxims to take the form of universal laws then it is crucial that future agents occupying my body cannot be identified with me prior to their taking on my volitions as their own.

Let's try and make this a little clearer. If I will to go to the dentist to get a cavity filled, this decision is going to be prospective because the action I undertake will need to be carried out during some period of time after the action has been willed. But the agents responsible for carrying out that action can't be the same as me *prior* to their following through on the same willed maxim since our being the same agent depends on being able to share in our wills. Thus when I intend to go to the dentist I have to will a maxim, giving my action a form such that it is available to another agent as standard that that

¹⁹ Ibid, 64

²⁰ Ibid

agent can either reject or conform to. If I will my maxim only for *myself*, then I can't truly will it because I would not be able to coherently intend it to be followed through over a period of time. I would not be willing it for the future agents occupying my body.

Actually, it's not just future agents occupying my body that I need to be concerned about, it's also myself. Korsgaard holds that willing maxims universally is not just necessary in order to establish the identity of agents across time, it is necessary in order to establish the unity of any given agent at one time.²¹ It is only through identifying with a law in the form of a maxim that allows me to constitute myself as an agent distinct from the impulses that would otherwise determine my behavior. The maxim not only secures the possibility for commitment given alterations in my incentives and motivations across time, it secures the possibility for commitment *now* in spite of competing incentives and motivations.

Korsgaard holds that willing maxims universally is not just necessary in order to guarantee the diachronic unity of agency, it is necessary in order to guarantee the synchronic unity of the agent as well. It is only through identifying with a law in the form a maxim that I can constitute myself as an agent distinct from the impulses that would otherwise determine my behavior. The maxim not only secures the possibility for commitment given alterations in my incentives and motivations across time, it secures the possibility for commitment *now* in spite of competing incentives and motivations. If I did not identify with a principle in contrast to my inclinations, then it seems like there are two possibilities. First, that I do not identify with a principle because I don't have one, in which case it seems like I am not operating reflectively at all and cannot be taken to be a

²¹ It's more accurate to say that willing maxims universally is necessary in order to establish the existence of an agent, since for Korsgaard agents constitute themselves by willing.

full-blooded human agent. For, in order to totally avoid willing a maxim I will have to simply and straightforwardly act by following some particular inclination since my inclinations are the only other available source of action. Second, that I treat the principle as distinct from me, in which case Korsgaard claims that I would effectively be turning it into "another force" equivalent to my competing incentives.²² This leaves me still to choose, among the options, how I identify.

What is important to take away from this discussion is that for Korsgaard it is a constraint on a willed maxim that it be willed universally. So, if my maxim does not universalize, it cannot coherently be the endorsement of a reason. Another way of putting the constraint is as follows: I can only will to do x for some reason if I could will that anyone do x for that reason. The next step in Korsgaard's argument for the publicity of reasons is to argue that since agents must constitute their identities according to principles of a universal form, the reasons I endorse when I will a maxim are public. We have to transition from a universalization of reasons that are private, where endorsing that toothaches are reasons for going to the dentist means when I have a toothache I have a reason to take myself to the dentist and when you have a toothache you have a reason to take yourself to the dentist, to a universalization of reasons that are public, where endorsing that toothaches are reasons for going to the dentist means that when you have a toothache it is also a reason for me to take *you* to the dentist and vice-versa²³.

There is an ambiguity here which we will have to concern ourselves with once we've reached the standpoint of public reasoning. Maxims may be a kind of commitment,

²² Korsgaard 2009, 75

²³ Of course, it doesn't have to be an overriding reason. Plus, considerations about how to divvy up labor among our public reasons may prevent counterintuitive results such as the suggestion that we handle any tasks on our own.

a law which governs how I behave if a particular situation occurs, but I may interact with those that endorse different maxims. What is ambiguous is whether or not the commitment that Korsgaard thinks is essential for maxims is a commitment *to the maxim* or a *commitment to act a certain way insofar as I endorse the maxim*. She seems to talk as if in the diachronic case, it is something like the former whereas in the case of interpersonal deliberations with other agents she can only mean the latter since I will have to come to a shared decision with the agents I interact with which takes into account my maxims and theirs, and which may involve the endorsement of some other maxim. The concern to be dealt with in Section VII is whether or not Korsgaard can treat the diachronic case distinctly from the interpersonal case.

Before I proceed to the next section, it is also worth commenting on the relationship between the example of the toothache and the Russian Nobleman. When Korsgaard argues for that volition requires determinate commitment, she seems to be addressing cases of temptation and attempting to show how succumbing to temptation involves some form of what she calls ‘particularistic willing’. Particularistic willing is the identification with a particular instance of inclination (for instance, fear) rather than a maxim with the form of a universal law. On the other hand, the Russian Nobleman case is not one of temptation, or at least need not be, as the shift in preferences between the nobleman’s younger and older self is presumed to be an enduring one and, further, the older Nobleman’s views about property are based not on a devotion to local impulse, but to a conscious self-preservation or greed. It is presumable that the older nobleman is working from a set of maxims that he endorses. However, the Russian Nobleman is a case in which it seems much more apparent that the past and future selves are different

people and so seems more intuitively to be a case of interpersonal rather than intrapersonal relations. In the final section of this paper, I will consider a case which does not involve particularistic willing but which involves an endorsement of a maxim which competes with the maxim willed by the past self, yet is not as intuitively interpersonal as the Russian Nobleman case. The question will be: whether or not Korsgaard's claims about the way in which volition commits us can have implications for cases *outside* of particularistic willing²⁴. Given her talk of 'good' and 'bad' reasons, it sounds as though she thinks that our maxims commit our future selves even when they do not will particularistically (since such willing does not produce genuine reasons). Examining such a case more closely, I will argue, shows that our volitions cannot do so if reasons are public.

In the next section I will consider Korsgaard's second step in her argument for the publicity of reasons as presented in the Chapter "Integrity and Interaction." This argument should help supplement her arguments that the categorical imperative is constitutive of action, making clear why the constraints of interaction demand a universalization of public reasons rather than a universalization of private reasons.

IV. The Argument from Deliberative Neutrality

The argument presented in "Practical Reason and Unity of the Will" constrains the form of the reasons that we can will according to the categorical imperative, namely

²⁴ It might be that this is all Korsgaard wants to show: that commitment just preempts particularistic willing. I don't think her arguments in *Self-Constitution* would bear out this reading, though it is possible. However, if particularistic willing is impossible (which Korsgaard claims) it is unclear how commitment could exclude it. Perhaps the fact that volitions are commitments is just a way of expressing that they can't be particularistic. In which case, it is unclear why Korsgaard cashes out the sense of commitment she is talking about in terms of maxims ruling out reasons. Only wills give reasons, and particularistic willing is impossible. Furthermore, Korsgaard suggests that the example bad reasons that she lists might be controversial, but never specifies the criteria for what might count as a bad reason. If reasons produced by particularistic wills were bad reasons, doing so would be straightforward.

that they take the form of a law which, in virtue of its general nature, can be taken on by other agents in a united will. This is not, however, a complete account of the publicity of reasons. As Korsgaard considers in "Integrity and Interaction", the universal form of maxims seems consistent with an account of reasons that is essentially private. This would imply that even when I take having a toothache to be a reason for me to go to the dentist, other people's toothaches aren't reasons for me to take any action (though I might will a maxim that would suggest that). In contrast, Korsgaard claims that if reasons are public then my willing our example maxim commits *me* willing that *you* should go to the dentist. If I take a toothache to be a reason to go the dentist, I take it to be a reason in a public manner: binding for everyone. The implication here is that I am bound to help you go to the dentist when you have a toothache. Of course, there are many, many other competing reasons which may override this reason, but your toothache is still a reason which makes demands of me as much as it makes demands of you.

Korsgaard's argument that reasons are public in this manner appeals to the fact that she takes human agents to constitute their own identities. Since human agents must reflectively endorse a principle in order to constitute their agency, there is no identity of the agent to speak of prior to the endorsement of some principle. Korsgaard thinks this fact means that there is no basis on which reasons could be grouped into 'mine' and 'yours' when engaging in practical deliberation. Here is the argument:

We *constitute* our own identities. So what counts as me, my incentives, my reasons, my identity, depends on rather than precedes, the kinds of choices I make. So I *can't* just decide I will base my choices only on *my own* reasons: because that category--the category of incentives that counts as mine and from which I construct "my reasons"--gets its ultimate shape from the choices that I make.²⁵

²⁵ Korsgaard 2009, 199

The idea appears to be something like this: maxims pick out what incentives count as reasons, but there is no class of incentives that would count as *mine* prior to establishing what maxims I endorse. Therefore, I can't pick out ones that will be 'mine' as opposed to 'yours'.

This seems convincing, but it doesn't seem to quite establish a way of counting reasons that is *universal*. For perhaps there is no set of reasons which is definitively first-personal from my perspective as an agent, but that doesn't prevent me from arbitrarily picking out certain incentives on the basis that they arise, say, from my particular body.

Korsgaard considers this concern and gives a response in a footnote:

Suppose I call my body 'Korsgaard' and I decide that I am going to attend only to the reasons arising directly from Korsgaard's thoughts and experiences, or something along those lines. That seems possible. But then I would have to be prepared to will it as a universal law that I should attend only to those reasons even if I turned out not to be Korsgaard.²⁶

The universal form of the laws that I will dictate that, while I can pick out a set of incentives that are restricted to particular embodied agents, the self which is to carry out the law I will cannot be arbitrarily restricted in this manner. In other words, the agent to which the incentives count as *reasons* can never be restricted to particular creatures, bodies or what-have-you. The agential self which carries out the maxim is identified with that maxim and so prior to its formation there is nothing to identify it with outside of the mere form of rational agency. Hence, whether or not Korsgaard is an agent in her own body, when she wills a maxim that takes the incentives originating in 'Korsgaard' as reasons, she is bound to take them as reasons.

²⁶ Ibid

It should be apparent that we are not, at this stage in Korsgaard's argument, to the publicity of reasons. Korsgaard wants the default to be that we treat others' reasons as normative for us, but we aren't there yet. Before we reach the final step of her argument for this conclusion, I want to consider an ambiguity on Korsgaard's account of agency: it is not at all clear how another's reasons ever could be normative for *me* if my identity is just constituted by the principle I endorse. I will only have the same reasons as another if we endorse the same principle. However, in virtue of endorsing the same principle, Korsgaard's commitment to the view that agents are only identified with the principles of action they endorse renders the result that we are no longer distinct agents. So, we ought to get clear on just what Korsgaard is going to want to argue for when she claims that others' reasons are binding on me by default. I will consider this issue in the following section, showing why Korsgaard must claim that the publicity of reasons can only result from a wholly shared volitional identity.

V. Pluralities of Agents and Pluralities of Principles

When considering what it means on Korsgaard's constitutive account of agency, to be bound by another's reasons we may want to ask another question: to what degree is it necessary to endorse the same principles as another agent in order to count as sharing reasons? For, in response to the all-or-nothing way in which we posed the difficulty, we might object something along the following lines: you and I can share in our reasons without being the same agent inasmuch as we mutually endorse a *particular* maxim, say the maxim of going to the dentist in case of a toothache. However, we also have many other maxims which we endorse which allow us to be distinct. While this will not get us to Korsgaard's claim that others reasons are normative to us by *default* it will allow us to

coherently conceive of others' reasons as normative for us in cases where we endorse some number of identical maxims.

I do not think this is a feasible response. The most significant problem with it is that it creates a problem which Korsgaard's account of agency does not have the resources to answer. Suppose that you and I endorse a plurality of principles, some of which overlap. What allows us to say that a certain set of those principles is *mine* and the other set is *yours*? We have to proceed very cautiously here for, on Korsgaard's account, there is no agential identity prior to endorsement. Agents are identified with the principles they endorse. One might have the kneejerk response: well, fine then, I am identified with this plurality and you are identified with that plurality. This still implies that there is some way to index an 'I' and a 'you' that underlies each plurality, attaches to them, to make them distinct. But as Korsgaard has to insist in her argument that reasons are public, we don't have the resources for that. The only way to identify agents is by their principles and prior to the act of endorsement there is no 'me' and 'you' to which principles could be divvied up.²⁷ Put another way, it doesn't seem as though on Korsgaard's account we can cogently talk about pluralities of principles all endorsed by one agent. In such a situation we would merely have a plurality of agents, each identified with each individual principle which has been endorsed. Thus a case of partial unity of will, a sort of overlap, will be impossible.

²⁷ This flies in the face of claims Korsgaard makes, especially in Chapter 10, which suggest that we can each have distinct practical identities. My arguments here are to show that, if she wants to hold this, she is going to have to appeal to resources beyond the bounds of what her account in *Self-Constitution* possesses and that she will have to be very careful to distinguish what allows our identities to be distinct in a way that does not provide a criteria for the private reasoner to then refuse her assertion that there are no criteria of identity that would demarcate private reasons from the deliberative standpoint.

Fortunately, Korsgaard is not stuck with an account of agency which will leave us all a large jumble of agents each operating on a distinct principle. Korsgaard takes the unity of agency to be central to her account and thus the obvious response for her is that when we will a plurality of maxims, we will them as a single maxim. That is to say, so long as we are taking them as normative for ourselves, the set of reasons they determine will all be normative for us simultaneously. If the agent is identified only with the principle it has endorsed, and we pick out the principle in terms of the action-reason pairs which it makes normative then the simultaneous normativity of a set of reasons which *could* be described in terms of a plurality of principles really is, from the agential standpoint, only *one* principle. This might seem like a view we would want to resist, since it seems natural to talk about a plurality of principles of action (after all, we don't cite every reason we've got when we explain why we've gone to see the dentist, just our toothache). All I can say is that Korsgaard will not want to resist it for fear of decomposing agency into a disunified jumbled of principles. It's also worth pointing out that our natural way of talking could just as easily treat the sorts of principles we are inclined to talk about as *really* being a plurality of principles. For instance, the maxim "I will go to the dentist when I have a toothache" could be decomposed into a plurality of principles: "I will go to the dentist when I have a toothache on Monday and I will go to the dentist when I have a toothache on Tuesday and etc..." Sometimes, this is the level at which we talk, especially if someone is puzzled as to why we went to the dentist at a particular time. There may still be something to be said about our common ways of describing action justifications at certain levels, but the issues here are too weighty for Korsgaard's account of agency for features of our language to be convincing without

further argument that our principles typically have certain 'natural sizes' aligning with common discourse.

All of this will lead us to a conclusion Korsgaard wants to endorse, though it is not quite the strongest form of the publicity of reasons. Korsgaard claims that interpersonal interactions, cases of cooperative agency, require the cooperating agents to treat each other's reasons as normative. This is because she takes cases of interaction to involve a unification of wills. If the unification of wills cannot be partial, as the above considerations seem to suggest, then unifying my will with someone else's means taking on board *all* of their reasons and vice-versa. There will be no way to pull apart just some principles and leave the rest. All of the other agent's reasons must come aboard. If diachronic action involves interaction with other agents, then it will be the case that we are also tasked with this project of unification. The difference is that my future selves don't exist yet, and so I cannot actively negotiate in order to reach some agreed, mediating principle that will govern our shared activity. The result is that self-governance will have to occur according to some principle that allows for the coherence of our actions across time in spite of our limited capacities to negotiate or to anticipate the reasons our future selves have. This principle will have to fall out of the constitutive norms of agency. Were it optional it would be possible that my later self would not endorse it, and then I would need to mediate between the principle of diachronic self-governance I endorse and the one my later self endorses and I could not appeal to the same policy. Whatever new policy I appeal to would have to be non-optional or the same problem would result. This is the conundrum I will consider in Section VII.

Now, it seems like the path to getting the strong sense in which reasons are public is becoming clearer. Korsgaard needs to establish two things: that *all* interaction must involve a *unification* of wills, second that the interactive stance is necessary in order to will. In the next section I will present Korsgaard's argument in favor of both these claims.

VI. The Final Step

Korsgaard's argument that the interaction of agents always involves the unification of wills rests on the claim that interaction "depends on the possibility of shared deliberation."²⁸ It is the possibility of shared deliberation that Korsgaard claims depends on the public status of reasons. Without the possibility that reasons can possess a normativity that "can extend across the boundaries between people"²⁹ interaction would be impossible.

Korsgaard gives a number of concrete examples to illustrate what she has in mind. The first example she gives is that of a student and teacher organizing a time to meet. The student and teacher have different constraints on when they could meet. For instance, the teacher can meet right after the class in which she teaches the student but the student has a class immediately afterwards and so cannot meet at that time. Since the act of meeting is a mutual one, Korsgaard wants to claim that teacher and student must deliberate together and so must share reasons.

Korsgaard rejects one possibility: one in which the teacher retorts to the student that she just needs to skip the class which interferes with the suggested meeting time. Korsgaard dismisses this case very quickly as a case in which interaction does not occur

²⁸ Korsgaard 2009, 191

²⁹ Ibid

at all. Certainly, the student and teacher are not deliberating together in this case. Further, they are only acting together inasmuch as the student acquiesces and takes the teachers' reasons to be normative, thereby abandoning treating class attendance as a reason not to meet at the time the teacher requires. When the student acquiesces there is still a unification of wills and still a sense in which mutual deliberation occurs, that deliberation can only result in the student abandoning a previously endorsed principle, however. Despite Korsgaard's brevity I think her evaluation is correct.

However, as Korsgaard indicates, this is a bit extreme of the teacher. There does seem to be a middle ground: "I could take your reasons into account, not as public reasons with normative implications for me, but as private ones with normative implications for you, implications which bear on my predictions of how you are going to act."³⁰ Korsgaard has a somewhat winding argument for why interactions cannot involve this sort of approach. I see no reason not to grant her claim, given that she takes interaction to be a kind of *cooperative action* and if she shows that cooperative action is the default for agency, then it won't matter what we call it: the default will be treating others' reasons as normative. However, I describe this middle-ground approach to negotiations because it is the last vestige of the private reasoner that Korsgaard considers as a live possibility. By the time Korsgaard begins her final argument for the publicity of reasons, she is considering the following option: the advocate of private reasons admits that, yes, when engaging in cooperation agents must mutually take their reasons as normative and thus share their reasons with one another but surely it's the case that I can still *choose* to cooperate or not. After all, then, it is a matter of choice whether or not I

³⁰ Ibid, 193

count others reasons as normative for me and so reasons aren't public: they are private but, perhaps, shareable when we decide to unite our wills in a cooperative enterprise.³¹

Korsgaard's argument against the optional nature of shared reasons rests on an argument that the act of identifying with a principle you have willed requires respecting one's own humanity and that 'humanity' is not a private thing such that one could respect it only in one's self, in some sort of private fashion. In arguing this, she appeals to issues in both the first and second steps of her argument: agential unity at a time requires a commitment to a principle that can be borne out across agents and that from the standpoint of one's bare agency, someone capable of reflectively endorsing principles or one's humanity, there is no way of picking out a particular 'me' in contrast to 'you' that would restrict my respect for humanity.

Let's consider the first part of this argument. Korsgaard has already argued that all action involves interaction. So she returns to the scenario of the Russian nobleman, who must either coerce his future self or cooperate with it, unifying his will with it, in order to carry out his maxim though only one of these options will allow the Russian nobleman to truly act in a diachronic manner. At the culmination of her argument, Korsgaard appeals to the diachronic case in order to show that interaction, which she has argued requires respect for humanity and therefore volitional unity, cannot be opted out of if we are to act at all. It is only through cooperating that we are capable of carrying out an action. Thus,

³¹ It is an interesting point that, given Korsgaard's view of agential identity, the fact that agents can unite wills and share reasons isn't actually a result of the public status of reasons. For, those reasons are only shared after the agents reflectively endorse the exact same principle, making them identical from an agential perspective. This still allows for the possibility that there can be reasons private to groups with united wills. Hopefully this makes clear why Korsgaard needs the stronger claim that reasons are normative by default for all agents. The reasons have to be shared *before* and not *after* deliberation. In other words, the demand that they be shared must be a constitutive feature of agency.

Korsgaard thinks that in order to act at all, we have to legislate *publicly* and endorse maxims in a way that treats their reasons as *normative* for other agents. If the Russian nobleman does not treat his reason as normative for his future self, as a law, then Korsgaard argues the Russian nobleman can't possibly be respecting his own humanity. This is because one's humanity is one's status as an agent capable of legislating about reasons in the first place. But this normativity must be public; otherwise the Russian nobleman cannot unify himself under it, whether at a time or across time. However, because the normativity of reasons is public it is not just the young Russian nobleman who legislates. If the older nobleman is rational, he also legislates with just as much normative authority. As a result, if the young nobleman is not committed to seeing his older self as rationally incompetent, he *must* respect the reasons he anticipates his older self to have. As Korsgaard puts it, he must either be at war with his older self or married to it.³²

Korsgaard is now making a different argument than the argument she initially made to support her claim that the categorical imperative is constitutive of agency, which instead leads to the conclusion that agency requires public reasons in the sense that those reasons are *binding* to later selves despite sharing the same basic appeal to the need for cooperation in diachronic agency. We might imagine the private reasoner resisting: certainly the young nobleman *could* treat his older selves' reasons as normative but that's up to the young nobleman. But the young nobleman doesn't *have* to. The twist in the diachronic case that Korsgaard adds is an appeal to humanity. Korsgaard wants to argue that the right the young nobleman has to form a will of his own depends on treating

³² Korsgaard 2009, 195

himself with respect, and treating himself with respect means treating his older self with respect.

So the situation is something like this: why not think that the private reasoner is committed to endorsing principles when acting but not to endorsing them publicly, that is as *normative* for her later selves? Well, she certainly must treat it as normative for her *now*. On what basis does she treat it as normative for her now? Because she respects her humanity, she treats herself as having a legitimate claim to legislating normativity into existence through volition. Two results are supposed to follow: most straightforwardly, inasmuch as she regards her future selves as also possessing humanity and thereby capable of volitional legislation, her respect for their humanity follows from a respect for her own humanity. Respecting hers but disrespecting theirs would be an inconsistency, and worse, incoherent according to Korsgaard. That would require distinguishing different 'possessors' of humanity, but from the deliberative standpoint Korsgaard has established that there is no 'I' or 'You' since no principles are as yet endorsed and so no agents are constituted. Thus, she must at least treat her future selves' reasons as normative for *her*. Less straightforwardly, since she recognizes that the humanity of her future selves warrants their legislation over her will she also recognizes that her humanity warrants her legislation over their will. However, the legislation of a will must ultimately be unified, or no action is possible. Thus, mutual recognition of reasons is necessary in order to arrive at a will that is not inconsistent with itself.

The result is that, as I have gestured at above, in a certain sense Korsgaard thinks all interaction is like marriage or friendship, at least inasmuch as it requires a certain volitional attitude of reciprocity: "friends *exchange* their private projects of pursuing their

own happiness, each undertaking to care for the other's happiness instead of his own."³³ Now, this is not a reciprocity that can be strictly understood as a kind of mutual self-effacement aiming at another's good because on Korsgaard's picture friendship is going to result in the formation of a *unified* will: "I pledge myself to pursue my friend's happiness, but her happiness in turn includes my own; she pledges herself to pursue my happiness, but mine now includes hers."³⁴ Preserved in the relationship is also the commitment to one's own happiness, albeit only through this reciprocal relationship. Korsgaard thinks, therefore, that "the exchange produces something new, a shared object, *our* happiness."³⁵ This is the structure of interaction for Korsgaard, not just for friendship or marriage, but for any interaction I undertake with another agent. Since it is central to Korsgaard's argument that interaction is *necessary*, this means that diachronic interaction is also something like friendship or marriage: it involves a mutual commitment to each other's projects³⁶. It is also a shared deliberation.

The structure of this account of volition thus underlies the fact that Korsgaard specifically insists in Chapter 4 (the first step of her argument) that maxims take the form of *provisionally* universal laws. A provisionally universal law is a law with a universal form but which incorporates exceptions into itself. To reiterate Korsgaard's example formulation of a provisionally universal law: "I commit myself to acting as this maxim

³³ Ibid, 188

³⁴ Ibid

³⁵ Ibid, 188-189

³⁶ It is natural to resist the idea that all interactions involve taking on others' projects in the same way as we do in friendship. While it may still be true in Korsgaard's account that facts about one's embodiment allow one to undertake the support of one's friends' projects in a more efficacious way, I think there are reasons to think that she really does think we need to be committed to the projects of anyone we interact with. The reasons to think this include the arguments I made in Section IV that the unification of a will cannot be partial as well as the fact that Korsgaard does explicitly claim that others reasons are always normative for us by default (this is, after all, how she characterizes their publicity). Korsgaard also explicitly states that she believes there is nothing in between war and marriage in cases of interaction. (Korsgaard 2009, 195)

specifies -- going to the dentist on the occasion of my appointment --... so long as I still have both the cavity and the appointment, and unless there is a good reason why not.³⁷

Korsgaard compares this to the provisional nature of our workaday understanding of physical laws³⁸, which we stipulate in a provisional manner and revise with a greater degree of specificity after observing exceptions³⁹. However, the provisional nature of maxims has its source in the structure of cooperative action. Since this structure involves shared deliberation, deliberation with shared reasons, it is best to understand the provisional status of maxims as accounting for potential changes in the normative landscape that would call for renegotiation of the maxim. With a physical law, future experiences might make us realize the inadequacy of our formulation as a description of the world or as a predictor of future events. With the laws of our will, what we might fail to anticipate are reasons we have yet to consider or encounter that would lead us to revise our deliberation.

This comes out most clearly in the case of diachronic action. If the binding force of an agent's maxim derives from a respect for humanity, that is to say the legislative authority present in any reflective agent, then when legislating it seems like an agent can't just dictate to her future selves how to act. She must also take into consideration her future selves' reasons, just as the reason she brings into being with the willing of her maxim must be taken on by future selves. This non-dictatorial, yet legislative attitude finds expression in the provisionality of maxims.

³⁷ Korsgaard 2006, 61

³⁸ As contrasted with a scientific understanding of physical laws.

³⁹ Korsgaard 2009, 74

Yet, there is a tension. Korsgaard wants the universality of maxims to explain the way in which our volitions are binding for our future selves. Inasmuch as maxims are laws, they keep us committed to a determinate principle of action and avoid leaving us a disorganized bundle of impulses. Hence the manner in which Korsgaard gives the provisionally universal form of willed maxims through the clause ‘unless there is a good reason why not.’ After all, “we are aware that certain unexpected circumstances could arise, circumstances that would give us *good* reason not to do it.”⁴⁰ Important here is the distinction between good and bad reasons not to follow through on the action. If there aren’t bad reasons to avoid going to the dentist, then the maxim no longer amounts to any sort of commitment. Korsgaard is insistent that the structure of volition is commitment, that commitment requires determinacy and that, therefore, I am not willing if “all I am doing on Monday when I commit myself to going to the dentist on Tuesday is committing myself to doing whatever I will decide to do on Tuesday regardless of my decision on Monday.”⁴¹ I am going to argue that Korsgaard cannot both think that we commit in this manner and that reasons are public in the sense that she has argued. This has more significant consequences than the construction of more theoretical scaffolding in order to distinguish how we reason diachronically, in an intrapersonal manner, rather than synchronically in an interpersonal manner. Korsgaard appeals to the diachronic case in her final argument for the publicity of reasons: our reasons must be public, binding universally, in order for us to undertake diachronic action. If the kind of interaction I have with future selves occupying my body is different in significant ways from the kind of interaction I have with other selves in other bodies then she can no longer

⁴⁰ Ibid, 77

⁴¹ Ibid, 78

straightforwardly appeal to the diachronic case in order to establish the publicity of reasons in the interpersonal case.

The idea, then, is that the agent initially willing the maxim anticipates that their future self might have good reason for acting otherwise, but still stipulates what would be a bad reason to opt out. If this is the picture of the provisionally universal form of the maxims we will and, therefore, of the way we unify our wills with others cooperatively there are complications for Korsgaard's account. In the next and final section I will explore these difficulties by looking at diachronic cases of cooperation. Before I begin I want to make clear what is at stake in the following section. The aim will be to understand in what sense, if any, Korsgaard can preserve the determinacy of particular commitments such as expressed in volitions to make visits to the dentist, while still arguing that we must reason publicly. Here is the difficulty we will be concerned with: Korsgaard argues that interaction requires the unification of the will with those we interact with, requiring a kind of shared deliberation; Korsgaard also argues that in order to act at all, we must interact with future selves occupying our body, unifying our wills with them; because we are neither omniscient nor prescient, we do not know what our future self's will would be; the only way we can secure the possibility of volitional unification, then, is either for the past self to let the future self decide or for the future self to let the past self decide; if the future self must let the past self decide, then it must be because of a constitutive norm of action; if it cannot be established that the past self has authority over the future self as a result of constitutive norms of action, then the past self must let the future self decide; the past self letting the future self decide requires an abandonment of the determinate commitment Korsgaard describes in Chapter IV and thus

requires rejecting the treatment of our past volitions as peremptorily binding. My main purpose is to work through a tension that ultimately amounts to an inconsistency in Korsgaard's argument in *Self-Constitution*, albeit one which I do not think ultimately prevents Korsgaard's account from being coherent. Since I will argue that it is not a constitutive norm of agency that our volitions be commitments, I will also have shown that Korsgaard can jettison those claims without undermining her account of agency. However, those that find Korsgaard's ability to account for something like the stability of intention important, especially as support for her account of volition, will find the arguments in Chapter IV thereby weakened.

VII. Anticipating Reasons

Consider Korsgaard's example of the student and teacher trying to coordinate a meeting. In the version where the teacher acts as a private reasoner, the teacher tells the student to simply skip the class which conflicts with the proposed meeting time. She is not treating the student's reason not to skip class as a reason when scheduling the meeting. Now let's put a different spin on the scenario: the teacher is not reasoning privately, and thus does treat *some* of the student's reasons as good ones. Suppose the student has to pick up her siblings from school, and so the teacher suggests a different time. The teacher takes *that* to be a good reason not to meet at the suggested time. However, the teacher does not consider competing classes to be good defeaters in their shared practical deliberations. She proposes a maxim, dictating when they will meet, and stipulates that competing classes are *bad reasons* to opt out of the action proposed by the maxim.

On the other hand, this scenario seems consistent with Korsgaard's description of the way we commit ourselves to a determinate course of action through volition. She seems to treat the intrapersonal diachronic case distinctly from the interpersonal case. Yet, to do so seems in conflict with how she describes the public nature of reasons: it is not up to the teacher to simply discount certain reasons the student has and likewise in the intrapersonal case of the Russian Nobleman. The teacher cannot disrespect the humanity in her student that allows her to legislate reasons nor can the young nobleman disregard the reasons of his older self. Presumably, the teacher/young nobleman can only discount the student's/old nobleman's reasons if they can't be legislated from the standpoint of humanity at all, that is to say: they are immoral⁴². However, Korsgaard seems to describe the stipulations made in maxims as to what might be good or bad reasons not to go to the dentist in this dictatorial manner: one stipulates that being afraid is a bad reason. It is clear that this can't be what one does, strictly speaking, for to do so would be to fail to reason publicly: if fear is a reason for my future self then I cannot fail to treat it as a reason. The examples Korsgaard gives of 'bad reasons' in the case of the trip to the dentist do not seem like they need to be immoral. Backing out of dentist trips due to anxiety or fear may produce a less pleasant world when universalized, but there doesn't appear to be any contradiction in willing such a maxim universally. Perhaps they would, but due the ambiguity with which Korsgaard presents examples it is still an open question whether the provisional universality of a maxim is constrained by morality or by something more

⁴² Korsgaard takes morality to fall out of the public status of reasons. Reasons are public because when I legislate (or when anyone legislates) the resulting reasons become normative for all humanity. But certain actions can't be willed for all humanity because they produce a contradiction in the will when legislated from this standpoint. I will be presuming that Korsgaard has successfully argued for public reasons and, therefore, morality. Since my concerns are with the internal coherence of her arguments, she or a defender could appeal to morality without begging the question or reasoning in a circular manner.

particular, like our practical identities. The latter certainly seems to fit one sense in which we often take ourselves to think of how we're committed. As a teacher, one might not think that a cold is sufficient reason not to be teaching, while perhaps an auditor might think that a cold is a good reason for not attending class. It is not necessarily a matter of morality for the teacher, though it might be. It could just be a part of her sense of commitment to her students.

One way to characterize this commitment is to think of it in terms of a certain kind of authority, albeit one that only results from an agreement to a particular way of conducting one's self. As I mentioned earlier when discussing the student/teacher negotiation over when to meet, the student might just unify wills with the teacher, thereby making it possible to interact by ceding authority to the teacher to make the decision of when to meet and regarding what reasons are good or bad ones for meeting at particular times. Of course, the teacher can't expect the student to do this unless she has a legitimate claim, in both their eyes, to this sort of authority. That is to say both must reflectively endorse the teacher's status as authoritative in this manner.

Appealing to a policy like this in the diachronic case seems like a way to preempt the concern about the tension the deliberating self is in when legislating for its future self. We might argue that in the case of diachronic action, the deliberating self has a claim to authority over the reasons of the future self. The argument might run something like this: we need commitment in order to have volition. We need volition in order to act. The only way to get commitment is if I cede authority to the volition of my past self. However, because we are first and foremost committed to public reasons and thereby committed to taking the reasons of all rational agents into consideration in our deliberations, in order to

successfully establish that this structure of diachronic mediation is warranted Korsgaard will need to argue that we are *constitutively* committed to ceding to the decisions of our past selves. Thus, if my future self decides not to go to the dentist because he is afraid, he is thereby making a mistake not just by my lights as the deliberating self, but by his own lights as an agent. If the argument is not made constitutively, then whether or not my future self decides to treat my volitions as authoritative is up to him and, as a result, I cannot genuinely commit him through my volitions.

By now we are familiar with the constitutive features of action according to Korsgaard: that we make ourselves into the unified causes of the ends we will. Failing to treat volitions of our past selves as committing us in the present must threaten the possibility of action, if commitment is to be possible. Here is an argument along those lines, given in the form of an example:

I find myself on Tuesday, about to go to the dentist and afraid to do so⁴³. I know that last week I decided I was going to go to see the dentist today. Now, I have all sorts of decisions I myself have made. For instance, I have decided that I want to see a rare live show of a foreign musical group, among others. I know that I'll be counting on my future selves to work with me in carrying out that action. Now, I imagine a future self in my shoes at the time of the show: it is a foggy night out and, despite it being a short drive that is relatively safe, that future self is afraid to drive. I know that *my* decision to see the show will only get carried out in such a circumstance if my future self treats my volition

⁴³ I want to make it clear that if the agent making the decision regarding whether or not to go to the dentist would not be trying to will particularistically should the decision be not to go because of fear. The agent would will the decision as a *maxim*, legislating fear as a genuine and generalizing reason not to go to the dentist. In other words, I am not presenting a case of what Korsgaard might consider temptation. I take my case to be more akin to the Russian Nobleman case (assuming the Russian Nobleman's future self can be considered rational).

as committing him, and therefore cedes to me that being afraid is a bad reason not to go to the show. But I am also in that situation with respect to my past self's volition to go to the dentist. I surely cannot make an exception of myself, for from the deliberative standpoint my past self and I are not distinct. If I want my future self to treat my volitions as committing him, I must treat my past self's volition as committing me as well. If I don't do so, I know that my capacity to act according to my ends will be jeopardized.

The demand I find myself confronted with in this example is parallel to the demand to unify myself by willing a maxim. Just as I am faced with a set of incentives which I must unite under a maxim in order to constitute myself as a unified whole, across time I may be confronted with a set of maxims which I must unite in order to constitute myself as a unified whole. I can't just identify with the maxim I have at any present moment, not now that I am required to deliberate publicly in a diachronic fashion. Instead, I need to mediate between all the various maxims my future and past selves might will and they must do likewise. I can do this only by forming some sort of principle of practical deliberation *either* to which we all agree or which is necessary in order to ensure our capacity to act. We will either need to adopt a principle which coherently governs the way our volitions are treated across time or be constitutively governed by one. It appears as though if I do not treat volitions as commissive, my capacity to act is threatened. So in the example given, the argument is that in order to satisfy the demands of the constitutive norms of agency volitions must be able to peremptorily commit (i.e. commit me against treating certain reasons as motivating where they conflict with the willed act).

Now, it is important to see that because of the nature of the diachronic case, if it turns out that the past self ceding to the future self is a viable option for conducting ourselves as agents, then it will be the *only* option. For, there is no in-between like in the case of an interpersonal negotiation between two people at a time where a back-and-forth can establish consensus without requiring any sort of asymmetry. The reason why it must either be the case that the future self gives say-so to the past self or vice-versa is because we are committed to treating our past and future selves' reasons as normative. Now, suppose I find myself deliberating about how to act and I must decide whether or not I'm making a commitment that involves overriding certain sorts of reasons my future self might have, which is to say I'm treating my volition as committing me even in the face of these reasons. Say I commit myself to going to the dentist even if I'm afraid. My will must be unified with my future self's will in order to act at a future date, because otherwise that will be some other agent and not *me*. So, if my future self refuses to treat my volition as committing him in this manner, then it will turn out that I have made it impossible to unify our wills, in the same way that the teacher being dismissive of the student's reasons prevents shared deliberation. Put another way, I cannot antecedently commit my future selves to uphold my commitments. This sort of self-authorization is circular unless it is the only option if I am to act: it must be authorized by the constitutive features of agency. Not only would an attempt to antecedently commit in this way be problematically circular, but by failing to accord my future self the same authority to legislate which I accord myself in virtue of our shared humanity, I actually undermine my own capacity to self-constitute. Thus, if volition can peremptorily commit in the face of

future reasons, it is an all-or-nothing affair (which is not to suggest that the commitments our volitions require of us are not defeasible).

We can draw out the dialectic of the situation by returning to the synchronic case. Suppose the teacher tries to apply a similar argument: "Look," she says "We have to come to *some* decision. Neither of us is budging so I'm just going to make a decision for us. There's really no choice, for we must figure out some time to meet." The immediate retort, available to the student, is that while it may be that someone has to make a decision, it surely doesn't have to be the teacher. "Why can't I be the one to decide?" she asks. We are certainly in this sort of situation in the diachronic case: someone is going to have to be the authority; either the past self will have to let the future self decide, or vice-versa. Is there a way to choose between the two? Otherwise, nothing is going to be decided, since they cannot reach out across time in order to negotiate actively.

Here is the answer that the past self will want to give as to why it is not the future self that can be the authority: suppose I let you, the future self, decide rather than vice-versa. Well, that would mean that I could not commit to anything. I might come to a final conclusion about what seems best for me to do at this particular moment but I could not really commit to any diachronically extended action, such as going to the dentist next Tuesday. Not only that, but neither could you, for you are also past self with respect to some future self. So you would really have to leave it to some future self to decide whether or not to carry out your decisions. And so the chain goes: you cede to him and he cedes to his future self and none of us form any commitments to do anything and so none of us really act. On the other hand, if you let *me* decide then a similar chain occurs, but one which strengthens our capacity to accomplish our goals. For you are a past self with

respect to some future self, and therefore in a position to commit that future self in the same way I commit you. Granted, many circumstances which we might not anticipate may lead us to reasonably abandon our commitment, but at least we would be committed to acting when things go as expected.

Now, the picture the past self paints of a world in which volitions commit us across time is certainly an appealing one in certain respects and may seem preferable to a world in which the deliberating self never commits its future selves. In this world, the past self leaves decisions about action at later times to the future self, which undermines the determinacy of our particular commitments, or so it would appear. The question of primary interest is whether or not this indeterminacy of commitment leads to some sort of incapacity to act effectively. The cost of self-governing in a manner which does not treat volitions as commissive is, presumably, that we lose the authority to set ends for ourselves in a manner that allows us to pursue them efficaciously as agents across time. The trade-off is that my past self has no say about what new ends I set at the present moment. In a sense, the tension between the two is something like this: if volitions are commissive, we give up our deliberative authority to *set* ends for ourselves as a result of our present evaluation of the reasons that motivate us. If volitions are not commissive, we find our capacity to pursue the ends we do set in virtue of being unable to *maintain* them in the face of redeliberation that we anticipate. We can here see that the issues of deliberative autonomy and agential efficacy begin to come apart in a way that Korsgaard does not consider. It is essential, however, to examine which is more fundamental to her account of agency and to establish how we mediate our deliberations with our future and past selves. I will now proceed to weigh the costs to the agential norms of autonomy and

efficacy in these two alternatives for diachronic mediation, beginning with the case in which our volitions peremptorily commit us in the manner Korsgaard suggests.

One thing to be concerned about if volitions peremptorily commit us is that what counts as a mistaken practical deliberation is determined only in terms of diachronic relations between agents. To illustrate this, I will suggest two scenarios: in one, I have tooth-pain, consider going to the dentist and decide that I am very afraid of dentists and that I'm not going to go. Now, as time passes this volition will commit my future selves. Suppose I even stipulate that increased pain would be a bad reason to go to the dentist. Later, as the pain becomes unbearable, I feel inclined to go to the dentist. Yet, I must accede to my past self's volition: I must not go to the dentist. However, in the second scenario, when I first deliberate I instead decide that the pain is a good reason to go to the dentist, and my fear a bad reason not to go. In this scenario, my later self is bound by my volition in the precise opposite way, even as the fear intensifies.⁴⁴ Outside of the authority of our past volitions, however, there are no grounds for being mistaken in our commitments to our maxim. If volitions commit us in this manner, it seems that they have a very powerful and largely arbitrary say in constituting what would be a mistaken practical judgment.

While most of us probably would find at least some of these decisions quite different from the ones we would be inclined to make, there are no constraints in the resources available to Korsgaard to exclude them. The priorities that our reasons have are

⁴⁴ One might wonder in what sense I could be the same person and be capable of, hypothetically, making completely opposite judgments about how I want to act. Of course, from the deliberative standpoint Korsgaard is committed to not distinguishing between agents at all and, further, to the freedom of the agent to self-constitute. One might also want to consider just how consistent people are about the way they weigh considerations from situation to situation.

totally dependent on the capacities of our deliberating self to stipulate them in its volition and on the stipulations it in fact makes. The result is that we can only be *procedurally* mistaken when we form practical judgments that conflict with judgments expressed in the volitions of our past selves. This is, of course, consistent with Korsgaard's account of the norms of agency if we suppose that volitions are commissive as a constitutive feature of agency: the norms that govern agency for Korsgaard are procedural or formal norms.

The difficulty with this picture of agency is the way in which it leaves the future self at the mercy of the past self. Sometimes we feel like our past self was in some way failing to appreciate the reasons we have, even if they anticipated the situation we find ourselves in. The perspective through which we are appreciating even the reasons our past self anticipated is one that self could not have occupied, so why should it be able to dismiss our perspective? If past volitions do not commit us, we have to give up our right to demand our future selves carry through the ends we set ourselves. But if we treat past volitions as able to commit us by preempting our reasons, we give up our right to exercise our capacities to reflectively endorse the courses of action we take⁴⁵. Perhaps the weirdness of this structure of diachronic rationality can be made clearer if look at a case where the past self has infinite cognitive capacities, such that they can consider every possible reason to abandon the commitment, and therefore establishes entirely in advance the contours of the commitment. The result would be that this past self could *definitively* and for all time determine how its future selves would be rationally bound to act, even

⁴⁵ Even if we want to insist on the right to redeliberate it doesn't mean we have to exercise it. The issue about whether it is resource-efficient or wise to always redeliberate has nothing to do with whether or not volitions have authority in the *face* of redeliberation.

though its commitment could have taken any number of mutually incompatible forms and therefore would be, in a certain sense, arbitrary.

The case of the cognitively infinite agent also draws out just how drastic the distinction between intrapersonal and interpersonal cases would have to be. After all, Korsgaard's arguments for the publicity of reason definitely commit her to the view that being able to anticipate and plan out our volitions for all future circumstances does not allow us to preempt the reasons of agents in other bodies. Yet, supposedly, the relationship I have to my future and past selves are the same as the relationships I have with other agents in other bodies, with the exception of the diachronic asymmetries confronting me in the former cases. It does not seem convincing that the structure of diachronic reasoning can warrant the sort of procedural constraints on diachronic practical rationality that have just been outlined. Indeed, something here seems to go against the whole spirit of the reasoning which was supposed justify the authority of our past volitions to commit us: it was as a part of a broader policy of mediation between past and future selves where each self traded full authority over its present course of action in order to secure the stability of the pursuit of the ends it sets across time. There must be more to the story of this mediation, then, since the case of the cognitively infinite deliberating self simply seems to annihilate any independence its future selves might have, thus making it an empty question whether or not those future selves ought to heed the deliberating selves' volitions in order to increase *their* efficacy. However, it seems uncertain that Korsgaard will have the resources to flesh out this story and how that would play out. There would need to be boundaries on the deliberating self's authority to deliberate for its future selves, but how those constraints would be established seems

difficult to ascertain. After all, they cannot be negotiated between selves. Furthermore, we cannot appeal to a principle in which the deliberating self must somehow respect the autonomy of the future self. For one thing, this leaves the question as to what degree the future self needs, minimally, to retain its autonomy. For another (and more problematically), the entire argument justifying the authority of our past volitions to commit us establishes that it is constitutive of the capacity of the future self to be an agent that it accept the commitments of its past selves. Thus, the commitments are not alien constraints on the autonomy of the future self to begin with.

Now that we have explored the case in which volitions peremptorily commit us, we can consider whether or not volition which does not commit us peremptorily undermines action. Is such an agent hindered in pursuing its ends? The answer is less clear when we consider that, bound by the need to reason publicly, the agent whose volitions do not peremptorily commit her sets her ends to be her best judgment about what she ought to pursue at any given time. Her long-term ends (insofar as she is committed to them) are indeterminate but she is not especially hindered in pursuing them and can do so until some deliberation leads her to believe that they need adjusting. Furthermore, they likely have some degree of persistence since we are unlikely, for reasons of economy, to redeliberate all our ends at all moments. Furthermore, our past volitions *still* carry weight as public reasons that will bind us in our deliberations. They just do not peremptorily bind us.

There does seem to be one thing missing if volitions do not commit us: a general policy for how to unite our maxims into a coherent and unified will. The demand to accede to past volitions at least gives us a concrete way to adjudicate between our past

and present judgments. Otherwise, how is one supposed to decide whether or not my present judgment, which may conflict with a past judgment, is mistaken or correct? Actually, this question is just as pertinent to the agent who is bound by past volitions. As a result of our finite cognitive capacities, our ability to commit through anticipation of future circumstances is very limited. In many cases, we're just going to have to fly by the seat of our pants⁴⁶. The difference between an agent governed by volitions that peremptorily bind and an agent governed by volitions that don't is that the latter simply flies by the seat of her pants all the time. However, we are no more undermined in our capacity to act diachronically when we encounter unanticipated circumstances than when we encounter anticipated ones.

Finally, approaching the relationship that I have to my past self's practical judgments in terms of correctness or incorrectness is simply mistaken itself, even if my practical judgments are the exact opposite of my past self's. It is mistaken in the same way that when we enter into shared deliberations with someone with competing practical judgments, we must assume that one of us is wrong. Inasmuch as Korsgaard is a constructivist, so long as those judgments are formed according to the constraints of agency, they are legitimate. It is clear that there is a great deal of wiggle-room to legislate in the sphere of the non-moral, such that two legitimate but competing judgments could occur. Indeed, on Korsgaard's account our practical judgments are acts of self-constitution in the face of the incentives and reasons we find ourselves faced with, not

⁴⁶ The exception, of course, would be if we were cognitively infinite and capable of planning for every instance. See above for discussion of such a case.

judgments about the reasons we have⁴⁷. When we encounter someone with competing judgments, we are both of us forced to find some way to mediate them and form a new judgment about how we are to act together while according each other's reasons weight. In other words, we are to self-constitute as a new agent, acting together toward whatever we decide is our common end. With that said, it's clear that if volition is non-commissive there is a principle guiding this mediating process. It is a kind of division of labor, akin to a relay race, in which each agent takes care of practical judgment and action at her time by taking into account the balance of reasons her past selves and circumstances have left her with. Yet, just as we can all effectively run a relay race together despite only running a portion of it individually, the agent whose volitions are non-commissive is capable of diachronic action as a whole in virtue of the fact that most fundamentally, every agent is committed to ceding to the next agent in line as time passes. The diachronic action to which such an agent is ultimately committed is just much more general than the particular actions she engages in at any given time.

If volitions are peremptorily binding, we are enabled as agents to be efficacious in the pursuit of particular ends. If volitions are not peremptorily binding, we are enabled as agents to exercise the legislative authority of our humanity unhindered. Actually, this evaluation of both cases is mistaken. Neither constitutive feature of agency is undermined in either case. If volitions are treated as commissive, then they are only commissive as a result of a constitutive need to endorse them as such, which means that it is constitutive of an act of reflective endorsement that it produce a volition which will commit our

⁴⁷ Thus, one could not argue that failing to treat volitions as commissive expresses an irrational distrust for the deliberative capacities of one's past self. Those deliberative capacities are only exercised in the act of self-constitution, which is precisely what the future self finds itself engaged in doing. Outside of the constitutive norms of agency, however, there are no standards of correctness for the way we self-constitute.

future selves. If volitions are treated as non-commissive, then I may freely pursue my particular ends, it is just that I do not have the authority to determine whether or not other agents, including my future selves, pursue those particular ends. My practical judgments are not voided, they are merely highly provisional, missing the direct stipulations that a peremptorily binding volition possesses. In many ways, the agent governed by commissive volitions and the one governed by non-commissive volitions are in the same boat, since they both must leave their commitments open to redeliberation in the face of encounters with other agents they may need to enter into interpersonal cooperation with or in the face of circumstances that they were not able to anticipate in their practical deliberations.

As I argued above, if it turns out that both commissive and non-commissive volition seem like viable options for diachronic self-governance, then it will turn out that commissive volition is never a viable option for diachronic self-governance since it is only such an option if it is constitutive (i.e. it is the only option). To reiterate, if commissive volition is not a necessary norm for agency then it may either be endorsed or not endorsed by later selves with whom I must cooperate in order to be able to act. It *is* a constitutive norm that I cooperate with them. Thus, if I insist on commissive volition and my later self disagrees, then I will have failed to cooperate and will undermine my own efficacy in diachronic action. The peculiar situation of separation across time in cases of diachronic action leads to the result that our volitions cannot take the form of commitment that Korsgaard describes in the first step of her argument for public reasons.

Conclusion

I have attempted to show that Korsgaard's commitment to the publicity of reasons is incompatible with her account of volition as determinate commitment. I think, at the least, I have shown that there are serious tensions between the two and that Korsgaard has not paid enough attention to the issues particular to diachronic agency. Perhaps this is because she does not think it necessary, since the unification of wills that constitutes our cooperation with other agents "cannot take place under the conditions of space and time."⁴⁸ Regardless, much elaboration is needed in order to make clear what the precise demands of public reasons are. This is crucial in the diachronic case, since the diachronic case is the case she ultimately appeals to in her argument to reasons are public. Furthermore, all our interactions with other agents in interpersonal cases will necessarily involve cooperation with future selves occupying the bodies of those involved, and so understanding how we reason practically in the diachronic case will affect how we reason in interpersonal cases.

The most fundamental point I have attempted to make is just this: if non-commissive volitions undermine our capacity to act, then we must be constitutively committed to commissive volitions and vice-versa. This is an important point when considering the issue, regardless of the success of my arguments that non-commissive volitions do not undermine our capacity to act, at least not more so than the many circumstances we are faced with which fall outside our anticipatory deliberations do so. At the very least, the stakes should be clear: Korsgaard needs to make further argument that commitment is constitutive of volition or abandon a model of diachronic agency that includes commitment.

⁴⁸ Korsgaard 2009, 190

Bibliography

Korsgaard, Christine. "Morality and the Logic of Caring." In *Taking Ourselves Seriously & Getting it Right*, by Harry Frankfurt, 55-76. Stanford: Stanford University Press, 2006.

—. *Self-Constitution*. New York: Oxford University Press, 2009.

—. *The Sources of Normativity*. New York: Cambridge University Press, 1996.