THE UNIVERSITY *of* EDINBURGH

Metaethical Constructivism and Treating Others as Ends

Ana Isabel Barandalla Ajona

Submitted for the degree of PhD by Research
The University of Edinburgh
2012

I have read and understood the University of Edinburgh guidelines on plagiarism. This thesis was composed by me and is entirely my own work except where I indicate otherwise by use of quotes and references. No part of it has been submitted for any other degree or professional qualification.

Ana Isabel Barandalla Ajona

Dedicado a María Esther

Contents

Abstract

Metaethical constructivism approaches metaethical questions from the perspective of the nature of normativity; and it approaches questions about the nature of normativity from the perspective of agency. According to constructivism, normativity originates in the agent. The agent gives herself laws, and these laws are normative because the agent has given them to herself.

Placing the agent as the source of normativity enables constructivism to answer metaphysical and epistemological questions about morality with ease. It also allows it to account for the relation between moral judgements and action.

But placing the agent as the source of normativity raises two questions. First, if the laws that the agent issues to herself are normative because she issues them to herself, what are the standards of correctness of those laws? Second, if the agent is her own source of normativity, how can she accommodate the normative status of others?

In this thesis I explore whether constructivism can answer those questions. In Chapter 1 I argue that the constructivist account of normativity is rich enough to answer the first question. From Chapter 2 onwards I argue that constructivism cannot answer the second question. I argue that its account of normativity requires that the agent does *not* accommodate the normative status of others.

Introduction

What is morality about, and why does it matter?[1] These are the leading questions of metaethics. They are big questions, as the driving intuitions of both our concept and practice of morality bear on all the major philosophical fields: metaphysics, philosophy of mind, epistemology, philosophy of language, ethics (responsibility, authority). To make the task manageable, philosophers work out an account of how those driving intuitions about morality impinge on a single given philosophical field, and then broaden that account to cover the other philosophical fields. The aim is to craft mutually consistent and complementary accounts of morality as it bears on the different philosophical fields. If they can manage that, then we will have a comprehensive answer to the leading metaethical questions. Needless to say, this is a difficult task, as demonstrated by the persistency and magnitude of the problems encountered by the different metaethical views. Let's look at the most common ones.

We can take as our starting point an issue in the philosophy of mind. What kind of mental state is instantiated by a moral judgement? When we make moral judgments and when we discuss those judgements, we distinctively think that our judgements might be correct or incorrect. According to the orthodoxy, only beliefs or belief-like states can be correct or incorrect. So, moral judgements must be beliefs.[2] Having proffered this answer as our starting point, we might draw the following implications. Since beliefs aim to capture what the world is like, then morality must be part of the world in the same way as the other things that we try to capture in our beliefs are part of the world. Since the things we try to capture in our beliefs are independent of our beliefs, moral truths must also be independent of our beliefs. That is why our beliefs about them might be wrong. And that is why moral truths are truths for everyone. This gives us a metaphysical realist picture of morality. Moral truths exist independently of how we try to capture them.

[1] Why does *morality* matter, that is, not why does *what morality is about* matter, although of course we can ask that question too.

[2] Or belief-like states. I take this as given from now on, unless otherwise stated.

But realism struggles to provide unproblematic accounts of the ontology of moral facts. As Hume pointed out, we don't perceive right and wrong in the same way as we perceive chairs, clouds, and trees (Hume 1975). If moral facts are not as the other facts that we form beliefs about, then, barring an independently motivated account of their ontological status, it looks like the thought that moral facts are 'out there' for anyone to see, only serves the function of supporting the idea that moral judgements are beliefs. But that is an illicit argumentative move. So we don't have good reason to accept that moral facts are independent of our judgements about them.

Hume's observation above also raises epistemological problems for the realist. Epistemologists can trace at least the contours of an account of our epistemic relation to things like chairs, clouds, and trees. And central to that account is our perception of those things. But if we don't perceive moral facts in the same way - if moral facts are not the same kind of facts as facts about chairs, clouds, and trees - it is not clear how we get to know things about them, or even form concepts of them.

In addition, the realist has a difficult task accounting for the motivational aspect of morality. We most closely associate the motivation to act with desires and conative states in general, not with cognitive states. So, although realism, by putting moral facts outside the activity of agency, does well at accounting for the assumptions implicit in moral judgement and discourse, it does less well at accounting for our epistemic and motivational relations to them.

With these considerations about motivation in mind, we might have provided a different answer to our initial question of what kind of mental state is a moral judgement. If we think that moral judgements are motivating - that they lead us to action - and think that the only motivating mental states are desires, or desire-like states, like sentiments,[3] then we might conclude that the mental states instantiated by moral judgements are desires. This is one of the main ideas of expressivism: that moral judgements are expressions of desires. The metaphysical implications of this would be that moral facts don't reside in the world independent of the agent's patterns of reaction to them. In light of this view, the epistemological questions become questions about self-knowledge for the expressivist. Although epistemological questions about self-knowledge face difficulties of conceptualisation, it is not mysterious that we should have self-knowledge - that we have

[3] From now on I shall talk about desires to include desire-like states, unless otherwise indicated.

an epistemic relation to ourselves. Expressivism is also well equipped to account for the phenomenology of moral judgements. Often moral judgements are accompanied by feelings and emotions. If moral judgements are expressions of desires it is unsurprising that we should experience emotions associated with frustration - anger, disappointment, sadness - when we judge something to be morally wrong, and elation when we judge something to be morally right.

However, although I introduced expressivism as an answer to the question of what kind of mental state is instantiated by moral judgements, expressivism doesn't do very well accounting for some of the intuitions implicit in the practice of moral judgment and moral discourse. Both when we form normative judgements and when we engage in normative discourse, we seek to adhere to rules of theoretical reasoning. Rules of theoretical reasoning consist in logical relations between thoughts, that is, between cognitive states. But desires are not cognitive states. As such, it would seem that logical constraints don't apply to their relations to each other. But if logical rules don't apply to the relations between our normative judgements, then it seems that we cannot form moral arguments. And if we cannot form moral arguments, then it seems that much of our moral practice is confused.[4] So, although expressivism provides convincing accounts of morality as it bears on motivation, epistemology, some aspects of the philosophy of mind, and metaphysics, it is less successful in its accounts of morality as it bears on other aspects of the philosophy of mind, and the philosophy of language.

Metaethical constructivism (henceforth 'constructivism') begins its approach to the leading metaethical questions by looking at the nature of normativity. To ask, Why does morality matter?, is to as why morality is normative, or whether morality is normative - depending on how we take the spirit of the question. The aspect of normativity which the constructivist focuses on most is how it is experienced by the agent.[5] That is, they look at what is for the agent, from the agent's own standpoint, to be under a normative claim.

The agent experiences being under a normative claim as being under a claim to *do* something. For example, the necessity to move out of the way when you see the truck trailer swerving your way (Street 2008, p.240), or the necessity to do what one is convinced

[4] The best known articulation of this worry trades as 'the Frege-Geach problem', in honour of Geach who raised the worry by applying an insight by Frege (van Roojen 2011, sec. 4.1).

[5] See especially (Bagnoli 2002), (Street 2008), (Christine M. Korsgaard 1996). See also (Bagnoli 2011, sec.7.1).

is the thing to be done as when we commonly say 'I *must* do this', or 'I *couldn't* do that' (Christine M. Korsgaard 1996; Velleman 2009; Bagnoli 2002).

This kind of claim is different from other claims, such as desires, on account of its origin. Its origin, according to the constructivist, is the agent herself. More specifically, it is an activity of the agent. Desires, by contrast, are not thought to not be properly the agent's own, because they are not thought to be activities of the agent. Rather, they are thought to be reactions by the agent to her circumstances. So, according to the constructivist, the distinctive nature and phenomenology of normativity as a claim in the will is given by its origin in the agent herself.

If the agent is the origin of normativity, then the metaphysical status of morality will be that of a 'construction' - a mind-dependent phenomenon we create and mould, and through which we mould our lives. Epistemic and motivational considerations cease to be mysterious: since it is all within the agent, epistemic and motivational connections are as straight forward as they can be.

But if the strength of constructivism comes from placing the agent at the origin of normativity, so do its weaknesses. The first concerns the constitutivist's ability to give a thorough account of the normativity of practical reason. It is not clear how normativity can arise from the agent unless the agent is normative. And it is not clear that the agent is normative. This gives rise to doubts that the constructivist can ultimately make good her account of the origin of normativity without either collapsing into expressivism, or falling back on realist assumptions. The second challenge concerns morality. Given the self-centred account of normativity provided by the constructivist, it is not clear how we can introduce morality into the picture without appeal to self-interest.

In this thesis I examine how and whether constructivism can meet those challenges. I shall proceed as follows. In Chapter 1 I present a generic account of constructivism, and consider whether it can meet the first challenge. I present that challenge as raised by David Enoch. Enoch argues that constructivism cannot account for the standards of correctness of action unless it presumes an implausibly permissive internalism about reasons. I argue that constructivism can answer Enoch's challenge by reaching out to Michael Smith's internalism. I will also argue that Smith's own account would benefit from this alliance.

In the reminder of the thesis I examine the question of whether constructivism can accommodate Kantian morality - a morality based on the idea of treating agents as ends.[6] To do this, I make Christine Korsgaard's work my focus. This is both because she has one of the most developed accounts of a constructivist morality, and because I believe that her overall constructivist account - not just her account of morality but her account of normativity - has been widely misunderstood.

In Chapter 2, I hone in on the challenge that morality poses to constructivism. I argue that the challenge is *prima facie* one of conceptual incompatibility. It looks like the constructivist's account of normativity presents a conceptual barrier to the requirements of a moral system. I then present a common interpretation of Korsgaard's attempt to meet the moral challenge - what I call 'the standard reading' of Korsgaard's work. Korsgaard's main device in her attempt to establish morality is what she calls 'the publicity of reasons'. According to the standard reading, Korsgaard's notion of public reasons is equivalent to Thomas Nagel's notion of agent-neutral reasons. Under this conception, not only does Korsgaard's notion of the publicity of reasons fail to meet the moral challenge, but the notion is itself riddled with equivocations and illicit inferences. I believe that the standard reading is mistaken, and that its mistake is based both on a misunderstanding about what Korsgaard is arguing *for*, and on a failure to appreciate the constructivist background against which she develops her thesis of the publicity of reasons.

In Chapter 3, I develop my own reading of Korsgaard's thesis of the publicity of reasons. This includes an extended presentation and reconstruction of Korsgaard's account of normativity. According to my reading, Korsgaard's thesis of the publicity of reasons attempts to capture and vindicate the idea that agents stand on a normative relation to each other.

In Chapter 4, I examine the way in which Korsgaard's thesis of the publicity of reasons would furnish an account of morality. I conclude that although it doesn't yield the unqualified account of morality Korsgaard might have sought, it does give us an acceptable account of morality. Specifically, I conclude that the account of morality we obtain is not one according to which necessarily all agents must be moral; but one according to which with the exception of exceedingly rare cases, all agents must be moral.

[6] Henceforth, by 'morality' I will mean Kantian morality unless otherwise indicated.

In Chapter 5, I return to the main question regarding constructivism and morality. This is whether morality is compatible with a constructivist account of normativity. At this point, of course, that question translates into whether Korsgaard's account of morality is compatible with her account of normativity. I argue that it isn't. My arguments here take a detour through Kant. Korsgaard's account is meant to provide an improvement on Kant's own moral theory. I argue that the way in which Korsgaard seeks to improve Kant's theory severs Kant's connection between the moral law and autonomy. Since Korsgaard's account of normativity is fundamentally based on Kant's account of autonomy, and her account of the publicity of reasons is an attempt to vindicate the moral law, Korsgaard's own account of the publicity of reasons is in opposition to her account of normativity.

In Chapter 6, I look to Stephen Darwall's account of the second-person standpoint (SPS). Darwall's SPS is of interest because it attempts to account for the way in which agents are normative to each other in much a similar way as Korsgaard's thesis of the publicity of reasons does. But Darwall argues that according to his account, the SPS uniquely facilitates the agent's autonomy. If Darwall is right, then given the similarities between the SPS and the publicity of reasons we might be able to incorporate into the thesis of the publicity of reasons the features that enable the SPS to accommodate the agent's autonomy. One of the main aspects of Darwall's account, however, is not compatible with a Korsgaardian constructivism. I argue that adapting that aspect to a Korsgaardian constructivism results in the SPS being unable to accommodate the agent's autonomy in the same way as the publicity of reasons can't.

In Chapter 7, I turn to Carla Bagnoli's work. Bagnoli proposes a relational account of autonomy as an alternative to Korsgaard's. The shortcomings which Bagnoli sees in Korsgaard's account are different from the ones I see, but her account of relational autonomy promises to address the shortcomings I see just the same. According to Bagnoli's account of relational autonomy it is intrinsic to one's autonomy that one sees others as ends. If this is so, then the expression of one's autonomy - i.e. treating oneself as an end - will involve treating others as ends. Since the problem I raised to Korsgaard's account was that one's autonomy rules out treating others as ends, Bagnoli's account of relational autonomy promises to address that problem directly.

I argue that replacing Korsgaard's account of autonomy with Bagnoli's would enable us to accommodate the requirement to treat others as ends. However, the cost would be giving up the ability to diagnose the wrong of actions which fail to express the agent's autonomy.

My conclusion will be that constructivism as presented here cannot combine a thoroughgoing normative account with a moral account based on treating others as ends.

- 1 -

Metaethical constructivism

1.1.1. A sketch of metaethical constructivism

As I said in the Introduction, constructivism approaches the leading metaethical questions - what is morality about, and why does it matter? - through the notion of normativity. More precisely, constructivists focus on what it is for the agent to be subject to a normative claim, and what follows from that.[1] Examples used to evoke the experience of being subject to a normative claim include the necessity to move out of the way when you see the truck trailer swerving your way (Street 2008, p.240), or the necessity to do what one is convinced is the thing to be done, as when we commonly say 'I *must* do this', or 'I *couldn't* do that' (Christine M. Korsgaard 1996; Velleman 2009; Bagnoli 2002).

But claims on the will we have aplenty, and no-one would wish to say that they are all normative. An urge to take revenge, to have a third serving of cake, to cheat, are claims that we feel distinctively and urgently enough, but they are not normative. If the constructivist focuses on the phenomenological aspect of normativity, she has to complement it with an account of what makes normative claims different from other claims on the will - different both in their phenomenology and in their practical status.

1.1.2. Valuing

According to constructivists, what makes normative claims different from other claims on the will is their origin. Normative claims, the constructivist proposes, originate in the agent herself. Specifically, they originate in the agent's activity of valuing.

[1] (Bagnoli 2002), (Street 2008), (Christine M. Korsgaard 1996). See also (Bagnoli 2011, sec.7.1).

The notion of the agent's valuing as the source of normativity has been given different characterisations by different philosophers. Some talk about valuing, or willing, or about taking as reasons or establishing normative connections with one's environment.[2] Each of these different characterisations will proffer a somewhat different conception of the phenomenon in question. In the next chapter I shall look closely at one of them, as the task I undertake there (looking at how constructivist tries to account for morality) requires attention to the minutiae. But my main purpose in the current chapter is to provide an outline of constructivism sufficient both to appreciate one of the main challenges it faces, and to appreciate how it can answer it. This purpose can be met by a broader outline of constructivism, so long as the relevant features are included in that outline, and even if that involves overlooking some important differences between the different accounts. With that in mind, I proceed to articulate what I take to be the relevant features of the phenomenon of valuing, and what, according to constructivism, follows from them.

The agent's valuing is supposed to involve more than attraction or aversion. It involves in some minimal sense approving, or disapproving, of the thing in question. To the extent that that is the case, valuing is an activity of the agent. Normativity is generated here, in the activity of valuing. Notice that locating the origin of normativity in an activity of the agent is important if the constructivist is not to lean on realist assumptions. This is why she locates normativity in valu*ing*, rather than value. Of course it is plausible to think that agency, and the agent, are manifested or realised in activity. When in the text I refer to agency and the agent as the sources of normativity, I shall do so in light of that consideration.

A feature of valuing important for the constructivist is that in valuing something, the agent sets restrictions on herself as to what other values she might hold. For example, if I value making my way to Rome tomorrow, and I know that to make my way to Rome I must buy a ticket today, in valuing making my way to Rome tomorrow, I am committing myself to valuing buying a ticket to Rome today.[3, 4] If in valuing something A the agent is setting

[2] (Street 2008; Street 2010; Bagnoli 2011; Bagnoli 2002; Velleman 2009; Christine M. Korsgaard 1996).

[3] This example is taken from Street 2008, p. 227. Street talks about 'entailments' arising from valuings, and she characterises valuings as 'taking as reason' (Street 2008).

[4] One could point out that it is possible to value going to Rome without having an intention to going there. In that case, we might think, valuing going to Rome does not commit me to valuing buying a ticket today. I think that there is a sense of 'valuing' in which this point is correct. However, at this early stage of the presentation it might be better to aim for outlines of the key concepts, and leave nuances for a more advanced staged of our discussion.

restrictions to rule out values which would go against her valuing A, we might put it that in valuing something A, the agent seeks to preserve her valuing A. If that is part of what it is to value A, then we can put it more abstractly: the commitment to preserving one's valuing of something is constitutive of valuing that thing.

But if two valuings rule each other out, the agent has to be able to decide which to keep and which to relinquish.[5] This brings us to another feature of valuings. This is that valuings are ordered. They are ordered in accordance to their relative importance to the agent. The relative importance of valuings is a result of both how much they matter to the agent, and of relations of dependency between them.

Examples of how much different valuings matter for the agent are common enough. I value having my dinner at 8.30 pm, and I value spending time with my friend, but I value spending time with my friend more than having my dinner at 8.30pm. An example of relations of dependency between valuings is the following. I value happiness, and I value money because it can make my life a happy one. In this case my valuing of money is dependent on my valuing of happiness.

These orderings help the agent solve conflicts between valuings. When two valuings rule against each other, the agent is to discard the less important one for the sake of the more important one. If some valuings are either incommensurate, or equally as important as each other, conflict between those valuings will be unresolvable or might only be resolvable by appeal to other valuings. However, it is possible that in some cases there just isn't a way to solve a conflict between valuings (Christine M. Korsgaard 1996, p.126).

This specification of the ordering of valuings tells us what valuings we may or may not have in relation to other valuings. But it is indeterminate on what valuings one might have *simpliciter*. This leads to problems of relativism. The threat of relativism would be averted if there were at least one thing that *must* matter to all agents - if there were one valuing which all agents must have.

At this point different strands of constructivism part ways. Humean constructivists hold that there isn't any one such valuing. They settle for the conclusion that it is possible for

[5] This includes cases where a valuing goes against its object, as in the case illustrated by the hedonist paradox. The hedonist paradox is that valuing one's pleasure most of all results is less pleasure than valuing other things.

agents to have values that are perfectly abhorrent to most of us. For example, they say that if Caligula's tastes were perfectly coherent within his valuings, there would be nothing wrong with him (Street 2010, pp.370–371; Lenman 2010). Kantian strands of constructivism, by contrast, argue that there is one such valuing which all agents must have. Korsgaard argues that is it valuing of oneself as an agent (Christine M. Korsgaard 1996; Christine M. Korsgaard 2009); Velleman argues that it is self-understanding (Velleman 2009). Yet a different question is whether we can obtain morality out of that necessary valuing. Here again authors differ. Korsgaard argues we can, whilst Velleman is less sure.

This chapter is not aimed to test the ability of constructivist to accommodate morality. I will touch on that topic peripherally later in the chapter, and I will take the question seriously from Chapter 2 onwards. Because of this, I shall ignore at present how different aspects of constructivism impinge on the possibility of a moral account. Instead, in this chapter I shall address a different challenge for constructivism: that of providing a thorough justification for the standards of correctness it sets. I believe that constructivism is best equipped to address that challenge when it holds that there is a specific valuing which all must have (C.f. Street 2010, p.374). With that in mind, I shall continue my outline of constructivism with the inclusion of a necessary valuing which all agents must have - something which necessarily matters to the agent. Once again, different constructivists will secure this necessary valuing in different ways. The main idea, however, boils down to a further aspect of what is constitutive of valuing. That is that in valuing something, one affirms one's own valuing of oneself as a valuer - as a creature to whom things matter.

1.1.3. The normativity of practical reason

Normativity concerns amongst other things actions, and actions issue from practical reason. To show how normativity generated by the agent's valuing gets carried over to actions, the constructivist conceives of practical reason in a distinctive way. To present this it will help to have in view the main elements of the notion of valuing I have outlined so far. These elements are: that normativity originates in the agent's act of valuing; that valuing is an activity of the agent; that in valuing something the agent is committed to preserving that valuing; that valuings are ordered in accordance to how important they are to the agent; and that in valuing something the agent affirms her own self-valuing.

The next step towards showing how normativity gets carried over to practical reason consists in a further observation about agents. This is that we must engage with our environment. We must engage with our environment not just in pursuit of our own material perdurance, but in pursuit of things that we value. And engaging with our environment is not simply a causal affair, but a heuristic and evaluative one (Bagnoli 2002). That is, when I am to set on my way to meet you for lunch, the way in which I engage with my environment will not be determined solely by causal considerations, but by evaluative ones. If we've agreed to meet at 1pm, I'll have to work out how long it takes me to get there, so I know when to set off. That is a casual consideration. (Although my wanting to be on time might not be.) But if I take one route rather than another because it is prettier, or if I wear my black dress instead of my green dress because I want to look nice for you, those choices are not led by causal considerations.

So our engagements with the world involve engaging with it in an evaluative manner. To direct the evaluative ways in which we engage with the world is to make decisions about how we engage with the world. And to engage with the world on the basis of decisions we have made is to act - to perform actions. Since we want to preserve the valuings more important for us, and in making decisions for action we commit ourselves to valuings (by developing new ones, or expressing pre-existing ones), we want for decisions for action to preserve our more important valuings.

In this way, if the fact that my health matters to me is in conflict with the fact that eating the whole chocolate cake also matters to me, and the way in which my health matters to me is more important to me than the way in which eating the whole cake matters to me, then I will seek to prevent myself from forming the decision to eat the whole cake.

What is important here is primarily not the fact that eating the whole cake would probably be detrimental to my health. What is important is that valuing my health and the value implicit in my decision to eat the whole cake are inconsistent with one another. That is what is important because, since I am involved in my values, in valuing things that I know to be inconsistent with each other I am positioning myself against myself. In other words, what is important about my decisions is not their expected consequences; what is important

is the normative economy created by the particular processes leading to my particular decisions.[6]

This takes us to another distinctive aspect of constructivism. My commitment to preserving my more important valuings is expressed in the process of forming my decision for action. Since the process of forming my decision for action is practical reason, my commitment to preserving my more important valuings is expressed in practical reason. Since the origin of normativity is in my valuings, practical reason obtains its normativity in virtue of being the way in which I express my commitment to preserve my valuings - a commitment which is constitutive of valuing, recall. And since my decisions for action are conclusions of my practical reasoning, my decisions obtain their normativity in virtue of being the conclusions of the expression of my commitment to preserve my valuings. In this picture, my decisions are my reasons.[7] And the process of reaching a decision is agency.

There are four aspects of the picture developed so far which merit flagging up. The first is that constructivism endorses an existence internalist view about reasons. Existence internalism about reasons is the view that for some consideration C to count as a reason for someone A, there has to be a connection between C and A's motivational set, where an agent's motivational set comprises all of the agent's motives, or, in our terms valuings (S. L. Darwall 1985, pp.54–55; B. Williams 1980). Existence internalism is a view about the ontology of reasons. Specifically, it says that reasons are constitutively linked to the agent's valuings. This aspect of constructivism will be relevant to my discussion of Enoch's objection to constructivism later in the chapter.

The second is the role which the constructivist assigns to reason. It is uncontroversial that practical reason yields decisions, or reasons, for action.[8] On many philosophical views the role of reason is seen as solely instrumental to forming a decision. That might be so

[6] The question of what to do with the consequences in theories of practical normativity which seem to focus exclusively on the process of decision making has been addressed by Barbara Herman (Herman 1992).

[7] This outline ignores the fact that we think of reasons as having relative normative valency to each other. That is, they can be overriding or overridden, prima facie, bona fide, conclusive, etc. In fact, all of those reasons will have been conclusive to some deliberative process or other, but they might change their normative valency in the context of other deliberative processes. In a way, then, a conclusive reason is ontologically and conceptually prior to all the other qualifications. As such, in discussing reasons I mean to discuss conclusive reasons *to the particular deliberative process from which they issue*, even if they would cease being conclusive in the context of another deliberative process.

[8] Many philosophers keep the notions of 'decision for action' and 'reason for action' separate. Constructivists are more likely to think that both notions point at the same thing, as I have intimated in the paragraph before last, in the main text.

regardless of what the content of that decision is supposed to be: whether it is to represent an independent rule about what to do; or what would secure the best balance between wellbeing and suffering; or what would preserve one's valuings best.

According to constructivism, in contrast, reason is not instrumental to finding out what would preserve one's valuings. Rather it is constitutive of finding out what would preserve one's valuings. The metaphor of 'finding out' here is meant to capture the dual role of reason as making something and discovering what it is being made at the same time. The idea is that you make your decision as you discover it in the sense that by conducting your deliberation as an expression of your commitment to preserving your valuing, you are preserving your valuing. But your deliberation involves forming beliefs about how your valuing is preserved, and your valuing is preserved by how you conduct your deliberation. This circularity is not vicious, it simply describes in very broad terms the reflexive nature of practical reasoning, a process where practical reason reaches an answer (a decision) by seeking to reach an answer, and represents its own progress in seeking to reach that answer.[9]

We find instances of making something as one finds out what it is (or, finding what something is as one makes it) in other areas. For example, thinking about how you feel makes a difference about how you feel. So, if I ask you how you feel, as you think about it you will be both contributing to how you feel and discovering how you feel. Granted that the analogy between this example and the decision making case is limited. Most notably, in this example there is such a thing as how you feel, only it might be altered to some extent by thinking about it.[10] However, to the extent that we can appreciate in this example the phenomenon of both creating and discovering something, the point of this example is fulfilled.

The third aspect to which I wish to bring attention follows from the second. That is that constructivism yields a cognitivist account of practical judgements. As I have just explained, in the process of deciding what to do the agent at once realises her commitment to preserve her valuings as she forms beliefs about whether electing this or that course of action would

[9] By 'seeking to reach that answer' I mean that answer *de dicto*, not *de re* - the agent has not set to herself what particular answer to reach prior to embarking on deliberation. To do that is to fail at one's practical reasoning in a specific way. That failing will be central to the running themes of the thesis from Chapter 5 onwards.

[10] Actually, I believe that under further reflection this does not pose a limit to the analogy. Discussing it would take me away from my main topic, though, so I shall leave it.

or would not preserve her valuings. If practical judgements instantiate cognitive states, it follows that one might be wrong about one's practical judgements. Constructivism accounts for the standards of correctness of practical judgement through its *constitutivism*: the view that the standards of action are found in what is constitutive of agency. This is the fourth aspect I wish to highlight, as follows.

According to constructivism, in valuing something we are committed to preserve it, and that commitment is expressed, or realised, in practical reason. In other words, practical reason is just the shape which the expression of our commitment to preserving our valuings takes. The process of ensuring that our valuings are preserved through action is the process of agency. That process consists in practical deliberation. So practical deliberation is *constitutive* of agency.

We have also seen that practical deliberation is the way in which the agent expresses, or realises, her commitment to preserving her valuings through action, and that a decision is the conclusion of that process. Accordingly, a decision will be correct or incorrect depending on whether it has been reached through correct practical reasoning - through the unencumbered expression of the agent's commitment to preserve her valuings through action. So, practical reasoning is also the *standard* of agency. So, practical reasoning is the *constitutive standard* of agency.[11]

It is worth emphasising the theoretical benefits of constitutivism and its central tool, the constitutive standards of agency, as these will be tested shortly. As I have explained above, it provides a formal reference for questions about the rightness or wrongness of actions. In addition, it provides an answer to sceptical questions about why one ought to abide by reason. One is to abide by reason because to abide by reason is to abide by one's own commitment to preserve one's valuings. Since it is constitutive of valuing that we are committed to preserving those valuings, in so long as one values anything, one has a reason to abide by the commitment to preserve one's valuings.

To the extent that constructivism is successful, it is so thanks to putting the agent at the origin of normativity. But that is where its challenges stem from. One of the main challenges it faces is to account for how agents can be the sources of value in the world,

[11] I believe it was Korsgaard who coined the phrase 'constitutive standards of agency' in her (Christine M. Korsgaard 2008).

without relying on realist assumptions, and without becoming a brand of expressivism. The other main challenge is morality. If the agent is her own source of value, how can others feature in the agent's normative landscape in a non-self-interested way?

I take on the second challenge from Chapter 2 onwards. In the reminder of this chapter, I address the first challenge as raised by David Enoch (Enoch 2006; Enoch 2011). Enoch argues that constructivism cannot provide a thorough account of normativity because it cannot fully account for the normativity of action unless it assumes an implausible form of internalism. I argue that Enoch's challenge can be met by constructivism by appealing to a well developed internalism such as Michael Smith's. I will also argue that Smith's broader account of which his internalism is part, would also profit from constructivism.

1.2.1. Enoch's challenge

In his 2006 and 2011, Enoch's criticisms against constructivism are directed to the constitutivist aspect of constructivism. We have seen how the constructivist builds constitutivism: the constitutive standards of agency forge a link between our valuings and our interactions with the world; their normative force resides in that connection; because their normative force is based on their connection to the agent's valuings, constitutive standards serve to answer any question of why one should follow them.

For example, let's agree that the constitutive standards of building a house include sealing the walls and ceilings so as to keep the weather out. If you are building a house and ask why you should seal the walls and ceilings so as to keep the weather out, the answer would be that, if you didn't do that, you would not be building a house (Korsgaard 2009, pp.27–30). You must follow the constitutive standards of whatever it is you are doing in order for you to be doing that thing.

For that answer to be a fit answer to the question of why you should follow the constitutive standards of whatever you're doing, it has to be the case that you value doing that thing. That is because it is your valuing of that thing that gives the constitutive standards of that thing their normativity. If constitutive standards can answer these kind of sceptical questions with regard to particular actions, they must also be able to answer the same kind of sceptical questions with regard to tougher sceptical questions about agency, questions

like, why follow the principles of reason at all? (Korsgaard 2009, p.29; Enoch 2006, p.169; Enoch 2011, p.208). The answer to any such questions would be that, without following those constitutive standards, one would not be engaged in agency. As before, for this answer to be satisfactory, it has to be the case that one values agency. We might put the answer in conditional form: if one values one's own agency, then one ought to follow the constitutive standards of agency.

Enoch's criticisms target two aspects of that answer. Firstly, he argues that the antecedent might not obtain: one might not value one's own agency. Secondly, he argues that even if the antecedent obtains, the consequent does not follow: valuing one's agency is not sufficient to give its constitutive standards normativity – it is not sufficient to make it the case that one ought to follow the constitutive standards of agency.

What is needed, according to Enoch, is a reason to value one's own agency. If there is a reason to value one's agency, then one will be required to be an agent; and if one is required to be an agent, then one will be required to follow the constitutive standards of agency. So, only if we can produce a reason for agency can we answer the sceptical question of why adhere to its constitutive standards (viz. by the principles of reason).

Enoch argues that the constructivist does not provide such a reason. From that he concludes that the constructivist must be relying on the view that valuing something is sufficient for it to be normative, or, as I shall be saying from now on, reason-giving. I will explain and address these objections in turn.

1.2.2. Caring about agency

Enoch's first objection is that one might not care about one's agency.[12] The objection is built on the presumed plausibility of the idea that someone simply does not care about agency. The idea of someone who simply does not care about agency denies the constitutivist her conclusion that one ought to follow the constitutive standards of agency. That is because

[12] I proceed as if the caring alluded to by Enoch was equivalent to the valuings I have been talking about, and to motives in the sense of items of one's motivational set. I shall use 'caring', 'valuing', and 'motives' interchangeably. If instead I took Enoch's 'caring' to refer to a more frivolous attitude, but that would build a less challenging and interesting objection to constitutivism.

the constitutive standards of things gain their normative force from the agent's valuing the thing those constitutive standards are constitutive standards of. So if one does not value one's agency, then one is not bound by the normativity of the constitutive standards of agency (Enoch 2011, p.209).

In raising this objection, Enoch is conceding to the constitutivist the existence internalism which the constitutivist endorses. We recall that, according to existence internalism, reasons are constitutively connected to the agent's valuings, or motivational set (on page 16). We recall too that one of the key theses argued for by the constitutivist is that a reason is a conclusion reached by the agent following a process which expresses the agent's commitment to preserving her valuings (on page 16). So, constitutivism endorses the connection between reasons and the agent's motivational set described by existence internalism.

In this light, Enoch's challenge consists in pointing out that, for all the constitutivist has said, one's motivational set might lack the motive to be an agent - that, for all the constitutivist has said, one might just not value being an agent. If one does not value one's agency, then one is not required to adhere to the standards of correctness of agency. If the standard of correctness of agency is practical reason, it would follow that one is not required to follow one's practical reason. This would deprive the constitutivist of the ability to answer the skeptic.

Enoch grants that the constitutivist might have recourse to the view that valuing one's agency is constitutive of being an agent (Enoch 2011, p.212). He is right in thinking that the constitutivist has recourse to that view and that it would be an effective tool against his objection. Agency is the process through which we regulate our interactions with the world in such a way that they preserve our valuings. Valuing things then is part of being an agent. We saw that part of what it is to value something is to be committed to preserve that valuing (on p. 16). So, being an agent involves the commitment to preserve one's valuings.

Since preserving one's valuings is what agency consists in, then if one values preserving one's valuings, one must value one's agency. So valuing agency is part of what it is to be an agent. In so much as the agent values anything, she values her being an agent. That answers Enoch's challenge by showing that, necessarily, every agent has the motive to be an agent within their motivational set.

To help reposition ourselves, recall the conditional relation advanced by the constitutivist and targeted by Enoch: if one values one's own agency, one is required to follow the constitutive standards of agency. Enoch's first objection is that the antecedent might not obtain. This would not render the conditional false, but it would render it useless for an ethical theory of action. In any case, I have argued that, according to the constitutivist, the antecedent of that conditional relation *must* obtain when it comes to agents. Valuing one's agency is constitutive of being an agent. This would restore the constitutivist conclusion that agents are required to follow the constitutive standards of agency, that is, the principles of reason.

But Enoch also raises objections to the conditional relation itself, and in doing so, he questions the internalist account of reasons implicit in the constitutivist picture. For Enoch's question now becomes whether valuing one's agency is sufficient to make the constitutive standards of agency normative. That is, whether valuing one's agency is sufficient to make it the case that one is required to follow its constitutive standards. This is a question about what items of one's motivational set qualify as reason-givers (when conjoined with the relevant beliefs or considerations).[13]

1.2.3. Caring about agency and internalism

Enoch's suggestion is that for a motive, for a valuing, to qualify as reason-giving it itself has to be justified. Suppose that I enjoy shooting birds for fun - shooting birds for fun is one of the things I value, and so one of the items in my motivational set. And suppose that there are better and worse ways of doing that shooting - that there are constitutive standards of shooting birds for fun. Enoch's suggestion is that we would not think that I am required to follow the constitutive standards of shooting birds for fun. We would not think that because we think that I am not justified in valuing shooting birds for fun. Enoch's point is that valuing something is not sufficient for one to be required to follow the constitutive standards of that thing. For one to be required to follow the constitutive standards of that thing, one must be justified in valuing that thing.

[13] I shall take this qualification as given in subsequent talk about motives (i.e. items in the agent's motivational set) providing reasons.

According to Enoch the constitutivist does not offer any support for the justification of the motive of agency. He infers that, if the constitutivist does not offer any argument for the justification of the motive of agency, it must be the case that the constitutivist endorses an existence internalism according to which membership of the agent's motivational set is sufficient for a motive to be reason-giving.

But as Enoch points out, that is too strong a brand of internalism (Enoch 2011, p.217). According to that brand of internalism, I *would* be required to follow the standards of correctness of shooting birds for fun if I valued shooting birds for fun. Moreover, I would be required to follow the standards of correctness of shooting birds for fun if I valued shooting birds for fun, even if I wished I didn't value shooting birds for fun (Enoch 2011, p.216).

A brand of internalism which claims that valuing something is sufficient for one to be required to follow the constitutive standards of that thing would yield the threat of unrestrained relativity and practical chaos. If Enoch's criticisms are correct, constructivism rests on an implausible brand of internalism about reasons.

However, Enoch is mistaken in thinking that the constitutivist needs that strong brand of internalism to uphold the conditional connection that if one values one's agency then one ought to adhere to the constitutive standards of agency. The constitutivist does not need to hold that *all and any* items from one's motivational set can provide a reason, in order to be able to maintain that one's motive for agency is reason-giving. All she needs to hold is that at least *some* items from one's motivational set are suitable reason-givers, and that valuing agency is one of the items suitable for reason-giving.

A type of internalism such as that developed by Michael Smith (Smith 1994), would aptly allow the constitutivist to meet that need by filling in the details of the existence internalism inherent in constructivism. I attempt to defend that claim in the next section. In the section after that, I will look at the way in which constructivism, for its part, helps overcome some of the shortcomings of Smith's wider position.

1.3.1. Smith's internalism as an asset for constructivism

According to Smith's version of internalism, the connection between judgment and motivation is a rational connection. More precisely, it holds that there is a rational

requirement for the agent to be motivated by her judgments about what she ought to do. Thus put, it might seem that Smith's is a *judgement* internalism. Since Enoch's objection was to the *existence* internalism implicit in constitutivism (p. 16 above) we might worry that Smith's internalism won't speak to Enoch's objection directly. To appreciate why this is a worry, it will help to look at the contrast between existence and judgment internalism in a bit more detail than we have so far.

1.3.2. Which internalism?

The distinction between the two types of internalism was articulated by Stephen Darwall (Darwall 1983, pp.54-55). According to Darwall, the defining idea of judgment internalism is that it is a necessary condition for an agent's *judging* that she ought to φ that she be disposed to φ. To put it in the terms we have been using, judgement internalism says that if an agent judges that she ought to φ, φing is connected to her valuings. In contrast, the defining idea of existence internalism is that it is a necessary condition *for something to be a reason* that it be connected to the agent's valuings - to the agent's motivational set.

Judgment internalism, then, tells us something about the nature of the mental phenomenon of making a practical judgement. That is that making a practical judgment has a constitutive connection to holding a given motive – another mental phenomenon. Existence internalism, in contrast, tells us something about reasons. Reasons – normative entities which might or might not be reducible to mental phenomena – have a constitutive connection with the agent's motives.[14]

To be sure, these two types of internalism are connected. Judgement internalism, as I have said, is concerned with practical judgment. But our concern with judgement arises because judgements typically contain, or purport to contain, reasons. The concern with judgement arises from that fact partly because of our interest in how to think about reasons, including their relation to the agent. And that is where existence internalism comes in. However, to the extent that the two internalisms are distinguishable from each other by making their

[14] Again, I mean to include within motives, desires, cares, valuings and values, and all other items that would compose a motivational set. This lack of discrimination does not matter for this discussion, given its scope, but in other contexts it would matter, for example in my initial exposition of constructivism in section 1.1.2. For salutary warnings on why interchanging values and valuings with desires might be problematic in some cases, see Street 2008 p.230-231.

focal point a mental phenomenon, in the case of judgement internalism, or the ontology of reasons, in the case of existence internalism, they have different toolkits with which to address different questions.

The question posed by Enoch is better answered by existence internalism, for his question expresses the concern that the ontological constraints needed for the constitutivist account to be plausible are too loose. That is, Enoch's question presses precisely on the heart of existence internalism as he thinks it is being used by the constitutivist. So existence internalism would be the type of internalism to address that challenge, for judgement internalism has nothing to say about the matter of the ontology of reasons. This is why it might be a worry that we think Smith's internalism is of the judgment type.

However, as we shall see presently, Smith's connection between judgement and motivation is mediated by a connection between reasons and motivation. In other words, he accounts for his judgment internalism by appeal to existence internalism. And there he will be concerned with what items from one's motivational set may or may not provide reasons. So, despite initial appearances, Smith's account is equipped to address Enoch's challenge.

1.3.3. Reason-giving motives

Indeed, Smith's internalism is based on his analysis of normative reasons.[15] Judgements about reasons, Smith tells us, comprise beliefs, or belief-like states, about what is desirable (Smith 1994, pp.137-148). So to say that the agent has a reason to φ, implies the judgment that it is desirable that the agent φ (Smith 1994, p.150). If it is desirable that the agent φ, then the agent would desire to φ if she were fully rational (Smith 1994, p.150). So, to judge that one has reason to φ, is to judge that one's φing is desirable. If one judges that one's φing is desirable, then one judges that one would desire to φ if one were fully rational (Smith 1994, p.150).

The notion of the fully rational agent employed by Smith, is based on that of Bernard Williams', as outlined in his 'Internal and External Reasons' (Williams 1980; Williams 1981).

[15] Smith differentiates between motivational reasons and normative reasons. Motivational reasons explain the agent's actions, whilst normative reasons explain them *and* justify them. Since my discussion will concern only normative reasons, that is what I will mean when I use 'reasons' and its connates.

That characterisation comprises the following conditions: the absence of false beliefs; the possession of all relevant true beliefs; and correct deliberation. Correct deliberation includes, as one would expect, logical sensitivities, but Smith adds the ability to create and abolish desires depending on whether they contribute towards instituting or maintaining coherence within the agent's motivational set (Smith 1994, pp.158–159). So, on Smith's account, to have a reason to φ is for φing to be what the agent would endorse in circumstances of full knowledge and flawless reasoning.

This supplies the first part of the answer to Enoch's challenge: that it is not the case that *any* item of one's motivational set may provide a reason. If we appeal to Smith's internalism, only those of the agent's motives which would be endorsed by a fully rational agent would be suitable candidates for providing reasons. The second part of the answer to Enoch's challenge is to show that the motive to be an agent - the agent's valuing of her own agency - would be supported by a fully rational being.

1.3.4. The motive for agency as reason-giving motive

According to the characterisation of the fully rational agent, such an agent would endorse the motive for agency only if, in full knowledge of the relevant circumstances and flawless reasoning, she found it to be at least consistent with the coherence of the actual agent's motivational set. If the motive for agency were at least consistent with the coherency of the motivational set, then the fully rational agent would discover that fact. So, our question is whether the motive for agency is at least consistent with the coherency of the agent's motivational set.

Recall what the function of agency is, according to the constructivist. The function of agency is to govern our interactions with the world in such a way as to preserve our valuings - our more important valuings, to be exact (Section 1.1.3. above). The motive to be an agent, then, is the motive to govern one's engagements with the world in a way that preserves those valuings. Since it is in the nature of valuings that we seek to preserve them, the motive for agency would be consistent with any and all motives. So, the motive for agency would be consistent with the coherency of any agent's motivational set. Therefore, the fully rational agent would endorse the motive for agency, and thus, the motive for agency is a reason-giving motive.

One might object to the idea that the motive for agency would be consistent with any motive. Motives with contents such as 'I value not being an agent', or 'I value not valuing anything' would not be consistent with agency. However, these valuings would not jeopardise the motive for agency on two counts. Firstly, these valuings are self-defeating. By being self-defeating they would ensure their own demise. Secondly, in being self-defeating, they are inconsistent with themselves, so the fully rational agent would not endorse them. Let me explain.

We said that it is in the nature of valuings that we seek to preserve them. In so far as these valuings – i.e., 'I value not being an agent', and 'I value not valuing anything' - are valuings, then, we seek to preserve them. However, their content requires that they not be preserved. This creates a situation in which whether the agent preserves them or not, those valuings will be defeated. If the agent preserves them, that will go against what their content requires. So, in preserving them the agent will be defeating them. If, on the other hand, the agent does not preserve them, then they won't be part of the agent's motivational set. And, if they are not part of the agent's motivational set, there won't be a question of whether they stand on any relation of consistency with the agent's other motives. In this sense, then, those valuings are inconsistent with themselves. As such the fully rational agent would not endorse them.

The possibility of those self-destructive valuings would not jeopardise the status of the motive of agency. The motive for agency remains consistent with all other motives of the agent, and so consistent with the coherency of the agent's motivational set. As such, it would be endorsed by the fully rational agent. In terms of Smith's internalism, that means that the motive for agency is a reason-giving motive, and in being a reason-giving motive its constitutive standards would be normative, that is to say, we would be required to follow them.

The answer to Enoch's challenge, then, is that it is not the case that the constitutivist can maintain the normativity of the constitutive standards of agency only by relying on an implausibly strong internalism. That implausibly strong internalism is one which claims that any and all items of one's motivational set - any and all of one's valuings - are reason-giving. I have argued that all the constitutivist needs is an internalism according to which only some of the items of one's motivational set are suitable reason-givers, *and* that one's

motive for agency falls into that category. I have argued that the constitutivist can craft that position for herself by appeal to Smith's internalism.

1.3.5. The requirement for *the motive for agency*

Through the above discussion I have attempted to answer Enoch's challenge as he presents it in his 2011. However, in his 2006, his challenge is stronger. His challenge to the constitutivist there is not only to show that we have a reason to have the motive for agency in the sense that we are *justified* in having that motive, but that we have a reason to have that motive in the sense that we are *required* to have it. He presents his challenge there through the idea of a 'shmagent'.

The shmagent is someone who is not an agent, who is perfectly happy being a shmagent, and who challenges us to give him a reason to be agent, or to value agency (Enoch 2006, p.179). Can the constitutivist meet that challenge? Can the constitutivist show that one is required to have the motive for agency? I think she can.

In the constructivist picture of which constitutivism is a part, for something to be required, is for it to be necessary to preserve one's more important valuings (Section 1.1.2. above). So, for the motive for agency to be required is for it to be necessary for preserving our more important valuings. Since it is in the nature of valuing something that we are committed to preserve it, trivially, valuing something involves the commitment to preserving that valuing. So, valuing agency entails being committed to preserve that valuing. In that case, if one values agency, then one is required to value agency. Since the motive for agency is constitutive of being an agent, then all agents are required to have the motive for being an agent.

If, on the other hand, one is not an agent, and thus, one does not have the motive for agency, then one is not required to have the motive for agency. But notice that if one is not an agent, one is not subject to any requirements. This is because requirements arise from our valuings on account of it being in their nature that we – agents – are committed to their preservation, and that whether we succeed in that is up to us. But this is what it is to be an agent: to have the ability to preserve our valuings through actions. If one is not an agent, then it is not the case that one has the ability to preserve one's valuings. Since the ability to

preserve our valuings is necessarily involved in the generation of requirements, if one does not have the ability to preserve one's valuings, then one is not subject to requirements. So, if one is not an agent, one is not subject to requirements of any kind, not even the requirement to be an agent.

We have an answer to Enoch's strongest challenge – the challenge to show that one is required to have the motive for agency, as presented via his shmagent. The answer is that, if one is an agent, then one is required to have the motive for agency; and if one is not an agent, then one is not required to have the motive for agency. Since the shmagent is not an agent, then he is not required to have the motive for agency - he is not required to value agency.

Having looked at how constructivism can benefit from adopting a well articulated internalism such as Smith's, I now explore how Smith's internalism would benefit from adopting constructivism.

1.4.1. Constructivism as an asset for Smith's internalism

One way in which constructivism can enhance Smith's internalism derives from the answer we gave to Enoch's objection. The answer to Enoch had two parts: first, showing that there are constraints as to what items of one's motivational set qualified for reason-giving status; second, showing that the motive for agency meets those constraints and thus qualifies for reasons-giving status. Part of the argument for the second part of the answer involved showing that the motive for agency is fundamental, necessary, and universal for all agents. It is fundamental because it underwrites the commitment to preserve all other motives, and it is necessary and universal because it is constitutive of being an agent. The existence of such a motive will enhance Smith's internalism on various fronts. Let's see why.

1.4.2. Universal convergence

Smith's internalism is developed as an answer not simply to the question of how judgment and action are connected, but more ambitiously, it is offered as an answer to the question of how to fit our intuitions about morality together, given that prima facie they appear to pull

us in different directions. These intuitions are: that our moral judgements express beliefs about what is morally right or wrong (call this the 'cognitive intuition'); that moral judgements are practical in their issue (call this the 'practical intuition'); and that moral judgements are objective – that what is morally right or wrong does not vary from person to person (call this the 'objectivity intuition') (Smith 1994, chap.1). The portrait of Smith's internalism I presented in Sections 1.3.1. to 1.3.3. above was a selective one, directed by my aim of showing how he forges a link between judgement and action, so I could apply it to the constitutivist picture. In other words, I constrained my presentation of his position to how through his internalism he attempts to link the cognitivist intuition with the practical intuition mentioned above. But his internalism is also supposed to satisfy the objectivity intuition.

Smith believes that his account satisfies the intuition about objectivity because he believes that his analysis of normative reasons yields a non-relative conception of reasons. That is, Smith believes that, according to his analysis of reasons, if it is the case that A has reason to φ in circumstances C, then it is the case that that anyone in C would also have reason to φ (Smith 1994, p.173). According to his analysis of reasons, an agent has a reason to φ only if the fully rational agent would desire that the agent φ. So, for Smith's analysis of normative reasons to yield reasons non-relative to particular agents, it will have to be the case that all fully rational versions of all agents would converge in what they would desire for each agent to do or desire. One would not have a reason to do anything unless the fully rational version of all agents would converge on the same content of the reason.

Despite the ambitiousness of that idea, Smith embraces it (Smith 1994, p.173). Despite the vast differences between the motivational sets of different agents, he remains optimistic that full rationality, including a process of systematic justification of our motives, would land us on the same judgments as to what is desirable, and, importantly, on which items of agents' motivational sets should act as anchors for the criterion of coherency of the motivational set (Smith 1994, p. 173).

To put that in words in line with earlier discussions: Smith thinks that the problem presented by the differences between the motivational sets of different agents for the attainment of universal convergence in judgements would be overcome by setting some universal motives which would ensure consistency across the different agents' motivational sets, and that those fundamental motives would be agreed upon amongst the agents

themselves from their current, disparate, motivational sets. Once we are on agreement on what are the fundamental motives, they can act as ultimate points of reference for our discussions about what we have reason to do.

But that just seems too ambitious. The possibility of *all* agents - not just actual, but possible agents - converging on what they consider desirable, seems too remote to be taken seriously. One of the ways in which this is brought out is by considering the temporal dimension of the formation of our motivational sets.

Many of the items constitutive of our motivational set are acquired before our rational capacities are fully developed. Inevitably, many of our motivational items will not have been subject to the rational scrutiny involved in deliberation which Smith supports. That is, many of the items of our motivational set will have come to form part of that set prior to our becoming full blown agents. As such, we won't have had the capability to implement a rational criterion for which motives we adopt or promote. When we become full blown agents and subject our motives to the requirement of consistency with the coherency of the motivational set, since the material or substantive outcome of the application of that criterion is determined by what motives there already are, the criterion of coherency will just as likely further entrench the differences in the components of different motivational sets, as bring them together.

The consequences of this are significant. If a criterion for a normative reason is too remote from the agent's grasp, then the agent would not be able to arrive at any normative reasons. If the agent cannot arrive at any normative reasons, then we are faced with a nihilist picture of agents' actual actions, for no-one would have a reason for their actions.

Constitutivism would help here. Since we have seen that there is at least one fundamental, necessary, and universal motive – namely, the motive for agency – we don't have to speculate that agents would be able to decide on which motive to assign that role. It is already there, for every agent. Furthermore, not only would that motive be present in the motivational sets of all agents, but it would be exercised in every action by agents. That motive then would serve to anchor everyone's motivational sets. The possibility of convergence of judgments about what is desirable in this picture would be greatly enhanced, for now there is at least one motive from which the different agents cannot depart.

1.4.3. The motive to be fully rational

The fundamental, necessary, and universal motive of agency would complement Smith's analysis of reasons in other respects too. Notice that Smith's analysis, in effect, works by appealing to a conception of the agent according to which agency includes the motive to be rational. That motive is crucial for Smith's account to work. It is so at two related levels. First of all, it is that motive that would explain why the agent embarks on her quest to find out what the fully rational agent would desire in the first place. Secondly, it is what would give Smith's conclusion its bite. Smith's conclusion is that if the agent fails to form the desire which she has determined she would have if fully rational, then, she is being irrational by her own lights (Smith 1994, p.177).

Before moving on to make my main point, I want to briefly point out that Smith is not entitled to that conclusion. It does not follow from my thinking that if I were *fully rational* I would φ, that if I don't φ I am *irrational*. What follows is that I am *not* fully rational. But being irrational is not the only way of being not fully rational. Another way of being not fully rational is being *less* than fully rational. This matters because Smith's conclusion is supposed to carry normative force. It is supposed to carry normative force so the agent can be criticised for not forming the desire in question, *and* so the agent can yet form that desire. The normative force is indicated by the term 'irrational'.

However, that normativity loses its footing when the conclusion is that the agent is less than fully rational. That is because 'less than rational' is only a descriptive expression. It wouldn't count as a failing for a less than fully rational agent to not do what a fully rational agent would do. This is why we don't think that the cat is at fault because it doesn't work out that, since I am the one who pays for its food, the least it could do is to go to the shop to get it itself. The cat's rational capacities are less than would allow for that level of reasoning.

Equally, the less than fully rational agent would not be at fault for not doing what the fully rational agent would do. The conclusion that one is less than rational would be an epistemic one: I discover, or confirm, something about myself, and that is that I am less than fully rational. This would allow us neither to criticise the agent, nor would it move the agent to form the desire in question. For Smith to retain the normativity of his conclusion, he needs

to argue that the agent who does not do as she would if fully rational, is non-fully rational in an irrational way, rather than in a less than rational way.

But let's suppose that he can make good the conclusion that one is irrational if one does not act in accordance with one's reasons. In this way, suppose that you are the agent in question. You judge that if you were fully rational you would be moved to φ, and at the same time, you are not moved to φ. You therefore conclude that you are irrational. But that conclusion is itself a belief. So, in itself it is not going to move you to φ. In line with internalism, for it to move you to φ, it will have to pair up with a given motive, in this case with the motive to be fully rational. So, the motive to be fully rational is essential to get Smith's account of deliberation started, that is, to get the agent searching for what the fully rational agent would endorse; and for his conclusion to carry the significance he seeks to establish, that is, for the belief that she is irrational to be motivating.

Smith's attempts at an argument for the existence of the motive to be fully rational, are minimal. His support for the existence of that motive boils down to the observation that that is what would explain the general phenomenon of agents adjusting their motivations to their beliefs about what is right (Smith 1994, pp.71-76). But the claim that it is agents' *beliefs* about what is right that agents' motivations adjust to, presupposes that that is what agents believe they would endorse if fully rational.

As mentioned above, for agents to believe that this is what they would endorse if fully rational, they must have embarked on a deliberative process in search of what they would endorse if fully rational. But if they have embarked on a practical deliberative process to ascertain what the fully rational agent would endorse, they will already have a motive to do as the fully rational agent would endorse. If they have a motive to do as the fully rational agent would endorse, and that motive prompts them to embark on a practical deliberative process to ascertain what that is, then, of course, the agent will adjust her motivation to the result of that search. That is because the motivation to do something as per one's decision is but the continuation of the motivation to embark on the deliberation which yields that decision. So, Smith's consideration in support of the idea that the motive for fully rational agency is constitutive of agents relies on a circular argument.

The idea that the motive for agency is constitutive of agency is one that the constructivist shares with Smith. But the constructivist, as we have seen, has a detailed argument for why we should think that the motive for agency is constitutive of the agent. The motive for

agency is the motive to preserve one's valuings in one's actions. And that motive comes with valuing anything at all (at least non-self-contradictory valuings).[16] And it is a starting premise of constructivism that an agent is a creature which values things. So Smith's attempt to establish morality would be enhanced by joining hands with constitutivism.

1.5.1. Summary

In this chapter I have presented a sketch of a generic constructivism. According to constructivism, normativity originates in the agent through the activity of valuing. In valuing something we are committed to preserving that valuing. Since in engaging with the world - in acting - we might preserve or jeopardise our valuings, we seek to preserve our valuings through action. That is, our actions are aimed to preserve our valuings. The expression of the commitment to preserve our valuings through action takes the form of practical reasoning. The process of governing our actions so as to preserve our valuings is the process of agency. Practical reasoning, then, is the constitutive standard of agency: whether an action is right or wrong is a matter of whether the agent chose that action through the full expression of her commitment to preserve her valuings through that action. This is constructivism's constitutivism.

Enoch challenges constructivism by arguing that its constitutivism does not have sufficient normative support to serve as the standard of correctness of actions. He argues that for constitutivism to be able to do that job, agency itself would have to be normative. But the only way in which agency could be normative is on the basis of an implausibly lax internalism - an internalism according to which anything we care about can be normative.

In a way, Enoch's criticism reveals more a puzzlement about the very idea of agency being the source of normativity, than a challenge to that idea. My answer to Enoch has involved spelling out some crucial details of the view that agency is the source of normativity. It has also involved appealing to a more developed internalism than the constructivist's own: Smith's. To think that agency is normative we don't need to think that anything we care about can be normative. We only need to think that some things that we care about can be normative, and that caring about agency is one of those things. Smith's internalism allows

[16] I presented that aspect of constructivism through the course of Section 1.3. above.

us to show that agency is one of the things which can be, and is, normative if we care about it.

I have also argued that Smith's own account would benefit from integrating with constructivism. Specifically, constructivism would enhance Smith's efforts to account for the objectivity of morality.

But this should not lead us to believe that the road from constructivism to morality is therefore clear. Constructivism faces obstacles peculiar to it, and not to Smith's account, at the time of trying to accommodate morality. Exploring these obstacles, as well as possible ways to overcome them will occupy the rest of this thesis.

- 2 -

Korsgaard's public reasons as agent-neutral reasons

2.1.1. Introduction

When crafting an account of the nature of morality philosophers aim to be able to accommodate the strongest intuitions implicit in the concept and the practice of morality. Our moral intuitions are not all in consonance with each other. One intuition is that the moral worth of an action is a matter of whether the world is a better or worse place for people as a consequence of that action.[1] This would be a utilitarian view. Another intuition is that agents are to be treated in certain ways despite the consequences of doing so. A Kantian would hold that view. The fact that these intuitions are not in consonance with each other is plain in ordinary situations. Shall I tell you the truth even though it will hurt you, or keep quiet so you stay happy? If we think that the consequences are what matters, we will think that I ought to keep quiet. If we think that what matters is treating persons in a certain way despite the consequences, we will think that I ought to tell you the truth.[2]

As I explained in Section 1.2.3. above, the constructivist evaluates actions not by appeal to their consequences, but by appeal to whether they have been chosen through the right procedures. The right procedures consist in the expression of the agent's commitment to preserve her valuings. Given the reflective nature of the process of expressing this commitment, we might put it that the constructivist evaluates actions by appeal to the way in which she relates to herself in the process of choosing her actions.

[1] I am leaving out the question of whether non-human animals can fit into a moral system.

[2] This is a broad sketch which I hope suffices to make the point. Authors siding with either intuition have sophisticated accounts which might lead them to produce the opposite answers to the ones I have suggested in the text. This goes beyond my remit here.

This suggests that constructivism would be more suited to moral intuitions which emphasise treating agents in certain ways, than to moral intuitions which emphasise the consequences of actions. Because of this, I shall make the Kantian intuition my lead into the question of whether constructivism can accommodate morality. That is, the question I pursue in the reminder of this dissertation is whether constructivism can accommodate a Kantian morality.

2.2.1. Constructivism and treating others as ends

According to the Kantian intuition agents are to be treated not just as having instrumental value, but as a having a value in themselves - as ends. However, although constructivism's accounts of normativity and of action suggest that it would be especially suited to accommodate this intuition, I argue that it is in conflict with one of the main tenets of constructivism, namely, that in valuing something we are committed to preserve that valuing through action (Section 1.1.3.).

For one to be an end is for one to have the status to set obligations, to set categorical constraints on what may or may not be done to oneself.[3] I am an end for you if you see me as something to which certain things may not be done regardless of how this constraint impinges on your or anyone else's interests. For example, you might see me as something to which you cannot lie, or injure, or kill, even if doing any of those things would further your ends, or other people's ends. For the present purposes it doesn't matter what the things that you may or may not do to me are. All that matters is that there are things that you may or may not do to me in virtue of how you see me. If such constraints arise from the way you see me, you see me as an end.

To be an end contrasts with being a means to ends. If you see me as a means to your own, or to another's ends, you will think that the constraints that apply to what things you may or may not do to me are conditional on meeting your ends. In this way you might not lie to me because you know you'd get caught and you won't be trusted again, and not being trusted will be detrimental to the pursuit of other ends of yours. Or you might not cause me injury because you love me, and seeing me suffer will make you suffer, which you don't

[3] For my current purposes I consider *things that must be done* as a subset of *things that may be done*. I will indicate when the modal difference between the two becomes relevant.

like. In both cases, the constraints on what you think you may do to me arise from your own interests. To be solely a means, then, is to be but an instrument to some end or another, and that negates one's status to set ends.

So, treating another as an end is to treat her as setting ends for you; treating her as a means solely is to treat her instrumentally to some other end; and according to our moral intuitions, we are required to treat others as ends. With this in mind, let us see why constructivism is at least prima facie inhospitable to the idea that one is to treat others as ends.

I explained in Sections 1.1.2. and 1.1.3. that, according to constructivism, in valuing something the agent is committed to preserving that valuing, and that actions are engagements with the world regulated by the agent to preserve her valuings. This suggests that we treat things in the world instrumentally to preserve our valuings. That suggestion seems to be confirmed by everyday experience.

For example, I value live music. One of the ways through which I preserve my valuing of live music is by attending live performances of music. My attending live performances involves multiple actions: I purchase a ticket, I get dressed in a certain way to attend the performance (I can't go in my pyjamas), I make my way to the concert hall, and I settle into my allocated seat. In each of these actions I treat the things I engage with (the money to purchase the tickets, the clothes I wear, the concert hall and the seats in it) instrumentally. So, reflection on our actual actions seems to confirm the constructivist tenet that we regulate our actions in order to preserve our valuings.

But other agents are also part of the world with which we engage. If it is the case that we treat things in our actions instrumentally because our actions are regulated to preserve our valuings, then, in as much as our actions involving other agents are also regulated to preserve our valuings, it would seem that we would treat other agents also instrumentally only. That is, if my actions involving dealing with the person from whom I purchase my ticket, the pedestrians in the street as I make my way to the concert hall, and the other members of the audience, are regulated to preserve my valuings, it would seem that I would treat all those people instrumentally only.

That conclusion obviously goes against the moral intuition that we are to treat others as ends. But it also goes against everyday experience. It goes against everyday experience in

that that is not how I conceive of my engagement with those people. Sure enough, I see their instrumental status, and that is reflected in my actions. Those people, just like any other being, have causal powers, and my making it the case that I am sitting in the concert hall enjoying the live performance involves engaging with the world causally. So, if I am to be an effective agent, I had better take into account the causal relevance of those people to my ends.

But it doesn't seem true to me, as the agent whose conceptions of her actions give her actions their identity, that the consideration of other people's causal powers is exhaustive of how those people feature in my own conception of my actions. When I speak to the ticket seller and I am polite to her, I don't conceive of my action as being motivated solely by my wish to get good service and my belief that being polite will elicit good service. If I thought that bullying her would result in better service, I still wouldn't bully her. I wouldn't do it because I see the ticket seller as someone who is not to be bullied regardless of how that impinges on my ends. In other words, I see the ticket seller as an end.

We might think that if the conclusion that we *treat* others only instrumentally follows from the idea that we seek to preserve our valuings through our actions, the way to avoid it would be to show that we *value* others.[4] The thought here would be that, if I value others, I would seek to preserve that valuing in my actions, and thus I wouldn't treat them solely as means. In this way we would accommodate our moral intuition that we are to treat others as ends. Since in valuing something we are committed to preserving that valuing, and other things being equal we are required to do what we are committed to doing, then if we value others we will be required to treat them as ends. It would also accommodate my own conception of my interaction with the ticket seller. I treat the ticket seller politely, even if I believe that bullying her would result in better service for me, because I value her more than I value getting a good service.

However, that strategy overestimates the theoretical work that valuing others can do. In fact, valuing others *simpliciter* can account neither for my conception of my interaction with the ticket seller, nor for the intuition that we are to treat others as ends.

[4] Showing that if we value others we will treat them as ends would not be enough to establish morality. To establish morality we would have to show that one is *required* to value others. However I argue that we can't show that valuing others leads us to treat them as ends. Because of this the additional question of whether we are required to value others in order to establish morality is moot.

As I explained in Section 1.1.2., our valuings are ordered according to how important they are to the agent. The relative importance of a valuing for the agent is the result of both how much it matters for the agent – how deeply she identifies with it - and whether it underwrites another valuing or is itself underwritten by another valuing.

One of the implications of that ordering is that it determines what valuings the agent is to pursue or not when they conflict. For example, I value drinking whisky, but I value having a clear head more than drinking whisky. So long as that ordering is in place, when facing a choice between drinking whisky and having a clear head, I am required to opt for having a clear head. Equally, I value being happy, and I value money because it enables me to be happy, so my valuing of happiness underwrites my valuing of money. So long as this is the case, if my pursuit of money conflicts with my happiness, I am required to pursue my happiness at the cost of pursuing money.

In the same way, if I value others *simpliciter* I shall be required to reflect that in my actions only if doing so doesn't undermine more important valuings of mine. But that is not what it is to treat another as an end. As I explained above, for me to treat another as an end is to treat them in certain ways regardless of whether doing so impinges on other valuings of mine. So, my valuing another *simpliciter* does not necessarily require that I am to treat the other as an end.

For me to be required to treat another as an end I have to value her *as an end*. But it is not clear that constructivism can allow for the idea that I value another as an end. Let's look at what position valuing others as ends would have to occupy within the agent's ordering of values.

I have said that to value someone as an end is to see them as placing ends for oneself - as setting categorical constraints on what one may or may not do to them, regardless of how that impinges on anyone's interests (on page 38). I have also said that it is in the nature of valuing that we are committed to preserving our more important valuings (Sections 1.1.2. and 1.1.3.). Unless the requirement of treating another as an end is to clash with the general requirement that we preserve our more important valuings, if the agent were to value others as ends, it would have to be the case that no other valuing of the agent is more important to the agent than her valuing of others. Since the relative importance of a valuing is the result of how much it matters to the agent - how closely the agent associates herself with it - and of whether it is underwritten by, or itself underwrites, another valuing, it

follows that for the agent to value others as ends there will have to be no valuing with which the agent associates herself more closely, and that valuing will have to not be underwritten by any other valuing.

But it is not clear that the agent's valuing of others can occupy this position within the agent's valuings. This is because given the nature of valuing, there seems to be another valuing which is more important for the agent than her valuing of others could ever be.

I have been saying that, for the constructivist, in valuing something the agent is committed to preserving that valuing. But notice that it is the *valuing* that the agent is committed to preserving, not the object of the valuing. The agent might seek to preserve the object of her valuing as a way of preserving her valuing of that object. For example, in my commitment to preserving my valuing of live music I seek to preserve live music. But I seek to preserve live music because that is required by my commitment to preserve *my valuing* of live music. In other words, in the constructivist picture, the agent remains the source of normativity even when she confers value on things. The value of the things on which she confers value remains contingent on the agent's conferring of that value - i.e. on the agent's valuing it. So, in all her valuings, the agent is committed to preserving her own valuing of herself as a valuer. This is why for the constructivist the agent affirms her own valuing of herself in valuing anything at all (Section 1.3.5.).

If the agent's valuing of herself is affirmed in all her valuing, then we might think that that valuing underwrites all other valuings of the agent, and that it is the valuing with which the agent associates herself more closely. That is, it seems that the agent values herself as an end, and that that must be so given the analysis of valuing employed by the constructivist. It is the reflective aspects of valuing as construed by the constructivist that make it the case that the agent must value herself as an end if she values anything at all.

It looks like, by definition, those reflective aspects of valuing can obtain only within the agent's own valuing, and not between the agent and others. So it looks like valuing others, however important that might be for the agent, cannot be as important as the agent's valuing of herself. If that is so, then the agent's valuing of others will necessarily be subordinate to the agent's valuing of herself. If that is so it follows that the agent cannot value others as ends. If the agent does not value others as ends, she cannot treat them as ends.

These considerations suggest that constructivism does not have conceptual room for morality because it cannot accommodate the idea that the agent is to treat others as ends. The challenge that morality poses to constructivism is that it looks like, according to constructivism, we can't be moral.

2.3.1 .Kant and treating others as ends

The problem of showing that the agent is to treat others as ends is not unique to constructivism. Kant also tried (Kant 1998)[5], and the difficulties raised to his arguments are in some ways similar to the difficulties encountered by constructivism.[6] Kant tried to capture the idea that the agent is to treat all agents as ends in his Formula of Humanity. His argument for the idea that the agent is to treat all agents as ends, then, is his argument for the Formula of Humanity.

Kant argued that since one does not think that the things one values are valuable in themselves, one must think of oneself as conferring value on those things. To think of oneself as conferring value is to think of oneself as a source of value. And to think of oneself as a source of value is to think of oneself as an end in itself, that is, as something which may not be used solely for the sake of something else. Since this is true of each person, the law that follows is 'So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.' (GW 4:428-429).

Kant's argument is controversial both in its premises and its logical structure, but we are interested only in the controversy surrounding its logical structure. That is because that is all we need to appreciate the similarities in the difficulties encountered by both Kant and the constructivists in showing that the agent is to treat others as ends.

The controversy pertaining to the logical structure of Kant's argument is that it seems to run as follows: since one must think of oneself as an end in itself, and everyone else must also think of themselves as ends in themselves, one must treat others as ends in themselves.

[5] From now on, references to Kant (1998) *Groundwork of the Metaphysics of Morals* will be given using the abbreviation GW.

[6] This is not a coincidence: Kant's arguments are precursors of constructivism.

The objection to that argument is that it commits a non-sequitur (Korsgaard 1996, p.134)[7]. The Formula of Humanity issues a dual instruction: (i) to treat oneself as an end; and, (ii) to treat others as end. But it looks like at most, all that the argument establishes is the first part of the conclusion - to treat oneself as an end.[8]

According to Kant, the requirements of morality are requirements of reason. So the obvious way to try to make Kant's argument work is to look for a rational requirement to link valuing oneself and valuing others.[9] That is, we might try to see whether valuing oneself creates a rational requirement to value others.

One of the rational requirements we might try out is consistency. The idea would be that if I value myself as an end, and I see that you are like me, consistency requires that I value you as an end too (and thus that I treat you as an end too). But it is not clear that consistency can do that job. It may be true that I see that you are like me in some important ways – I see that you must value yourself as I must value myself – but we are also different in important ways: I value myself because my conferring value on the world commits me to it; and you value yourself because your conferring value on the world commits you to. Crucially, my conferring value on the world does not commit me to value you. So the consistency requirement doesn't apply. Valuing myself as an end, and valuing you as an end are too different from each other for there to be a consistency requirement from one to the other.

Self-interest is another rational requirement which we might consider as a possible way of linking valuing oneself to valuing others. The idea here would be that it is in one's self-interest to value others. However, the idea of valuing another as end out of self-interest is an oxymoron. To treat another in any way out of self-interest is to treat them as means to your ends, whilst to treat another as an end is to treat them as setting ends for you regardless of your self-interests.

To be sure, I can treat you both as a means and as an end. If you purchase something from my shop and I give you the correct change because I judge that that is the right thing to do, I treat you both as a means and as an end. I treat you as a means to my livelihood – I take

[7] From now on, my references to Korsgaard's (1996) *The Sources of Normativity* will be given through the abbreviation SN.

[8] Some authors, eg. Darwall, think that Kant's argument does not even establish that one is to treat oneself as an end (S. L. Darwall 2006; S. L. Darwall 2009).

[9] From now on, and unless otherwise stated, by 'valuing' others or oneself, I mean valuing as ends.

your payment; and I treat you as an end by giving you the right change because that it is the right thing to do – because the money surplus to the price of your purchase belongs to you. But I can't give you the right change because it is the right thing to do out of self-interest. If I give you the right change out of self-interest, I am doing so, by definition, with a view towards forwarding my interests. In contrast, to give you the right change because it is the right thing to do is to do so *despite* my own interests. So I can treat you as an end *and* instrumentally, but I can't treat you as an end instrumentally (or instrumentally as an end).

Kant's argument for the Formula of Humanity bears a broad structural similarity to constructivism: they both show that the agent is to value herself as an end; and they both seem to face a conceptual impasse to show that the agent is to value others as ends.

2.4.1. Korsgaard and morality

Christine Korsgaard sets out to establish a constructivist morality in a way that, if successful, it would also vindicate Kant's argument for the Formula of Humanity. Her overall strategy is to argue that valuing oneself 'somehow implies, entails, or involves' valuing others (SN 132). If the constructivist is right that one necessarily values oneself, it will follow that one necessarily values others as ends. Since according to constructivism one is to preserve one's more important valuings through action, and for one to value someone as an end is for one to have no more important valuings than valuing that someone, then, if Korsgaard establishes that one necessarily values others as ends, she will establish that one is required to value others as ends.

That will give us a constructivist morality. It will also give us the missing link in the Formula of Humanity between the double instruction to treat oneself as an end, and to treat others as ends. If I am to treat myself as an end because I value myself as an end, then if I value others as ends, I will be required to treat others as ends. If valuing myself as an end involves valuing others as ends, and I necessarily value myself as an end, then I necessarily value others as ends. Therefore, if I necessarily value myself as an end, I am to treat others as ends.

To show that valuing oneself involves valuing others, Korsgaard develops a revisionary account of reasons. According to Korsgaard the main obstacle to see that valuing oneself

involves valuing others - the main obstacle to appreciate the sound structure of Kant's argument for his Formula of Humanity - is a private conception of reasons (SN 132-135). Under that conception, reasons are private in the sense that their normativity is confined to individual agents. My reasons are reasons for me only, and yours for you only. I might have a reason to take your reasons into account, but, on this view, having a reason to take your reasons into account is no different from having a reason to take the weather into account if it affects my plans. They have practical relevance to me, but only instrumental - they don't *set ends* for me, i.e. they are not normative to me in themselves (SN 132-134).

By way of illustration, we might think of Humean reasons as an example of private reasons. In very broad terms, the Humean thinks that, since reasons have to be motivating, and that only desires (broadly speaking) are capable of motivating, reasons must be desires. In this picture, since your desires can only be yours - just as your pain can only be yours, and your pleasure can only be yours - so can your reasons only be yours. To be sure, your reasons might *cause* me to have certain reasons, just as your pain might cause me pain. But these remain yours only, and mine only, respectively.

If this is how we think about reasons we find Kant's argument fallacious because it can't tell us why other's reasons to treat themselves as ends give one a reason to also treat them as ends. Furthermore, even if there was a reason why others' reasons to treat themselves as ends give one a reason to also treat them as ends, that reason would be hypothetical, as, on this view others' reasons have only instrumental relevance to oneself. We think that the requirements of morality are categorical. So, under this conception of reasons, not only is Kant's argument invalid, if we fixed it, it would not give us morality.

The right conception of reasons, according to Korsgaard, is precisely a conception in which reasons are public – public in the sense that their normativity applies to other agents. Under this conception, we find that valuing oneself involves valuing others. Once we see that, we also see that the argument for the Formula of Humanity is valid.

What Korsgaard means by 'public reasons' is not a straightforward matter. The most she says about it by way of characterisation is in a footnote. There she says that what she is calling 'private' and 'public' reasons 'are roughly what in contemporary jargon are called "agent-relative" and "agent-neutral" reasons.' (SN 133 n.3). Not only does that leave the reader perplexed about how much distance is covered by the 'roughly' between her conception of public reasons and agent-neutral reasons, but there is a plethora of senses in

which reasons might be said to be agent-relative and agent-neutral in contemporary jargon (Ridge 2005a).

In another article (Korsgaard 1993) Korsgaard discusses the distinction between agent-relative/neutral reasons most associated with Thomas Nagel. According to Nagel's distinction, a reason is agent-relative if its articulation makes an essential reference to the person for whom it is a reason; if it doesn't the reason is agent-neutral (Nagel 1986, pp.152–153). I take it that it is in consequence of this that several commentators have read Korsgaard's public reasons to be equivalent to Nagel's agent-neutral reasons. This is what I call the 'standard reading' of Korsgaard's discussion of public reasons. It has been adopted by authors such as Joshua Gert, Alan Gibbard, Michael Ridge, and R. Jay Wallace (Gert 2002; Gibbard 1999; Ridge 2005b; Wallace 2009).

I believe, and will argue later, that the standard reading of Korsgaard's public reasons is mistaken because it fails to take into account both the nature of the connection between valuing oneself and valuing others which Korsgaard's seeks to establish, and her account of normativity. However, since the standard reading has been endorsed by such illustrious commentators, and it follows so naturally from what little Korsgaard says about her conception of public reasons, it will be worthwhile to present it here. The merit of doing so is to show how badly Korsgaard's account fares under that reading. So badly that it should prompt us to look for an alternative reading. I shall continue on to provide such an alternative reading in Chapter 3.

2.5.1. Public reasons: the standard reading

Korsgaard's discussion of the thesis of public reasons is divided around two main arguments: an argument making use of Wittgenstein's argument for the publicity of language; and an argument about the permeability of one's consciousness to another's reasons. I take them in turn.

2.5.2. Public language and public reasons

Wittgenstein gave us the idea that natural languages are essentially public. The argument for that idea was centred on the thought that linguistic meaning is largely constituted by rules, and that rules can only be rules if someone requires that they be held to.[10] This aspect of meaning - its being largely constituted by rules - is the focus of Korsgaard's attention, specifically, the relational aspect of those rules. She formulates Wittgenstein's position thus,

> to say that X means Y is to say that one ought to take X for Y; and this requires two, a legislator to lay it down that one must take X for Y, and a citizen to obey (SN 137).

Gert takes it that the relation to which Korsgaard is drawing our attention there, is that between *words* and *rules* (Gert 2002, p.311). That relation is such that, for words to have meanings, they have to be under the governance of a rule stipulating what the meaning of those words is; and for a word to be used meaningfully (i.e. properly) is for its use to comply with the relevant rule. For example, for the sound <*window*> to be meaningful, there needs to be a rule which stipulates what its meaning is, and that sound has to be used to refer to what the rule says it is to refer to – in this case, to a window.[11] If there is no rule, the sound could be used willy-nilly, which is to say, it would be meaningless; and, if there is a rule but the sound is used to refer to something other than a window then the sound would also be said to be meaningless.[12] This, then, is the relational aspect of the normativity of meaning which, according to Gert, Korsgaard seeks to highlight in the above quote.

Gert further notes that that relation between words and their rules makes meanings public in two intertwined ways: whether a word meets the standard set by the rule or not, will be publicly observable; and words can be taught (Gert 2002, p.311). In this way, if I say, 'Look, Andrew is coming through the window', as I point at Andrew coming through the door, anyone who knows the rules applying to the words I have used can tell that I am misusing the word 'window'. And I can be taught what the meaning of 'window' is: you can give me a definition of the word, or point at the window.

[10] The nature of the rules of language is a disputed matter into which I do not wish to enter. The only commitment to which I subscribe is that they are normative (as opposed to, say, merely descriptive) - that is why Korsgaard uses them as analogous to reasons.

[11] I'm being loose about issues of meaning and reference here, but that is of no pertinence to my point.

[12] I am ignoring metaphorical uses of language.

Having laid out the way in which the relational structure of the normativity of meaning makes meaning public, Korsgaard draws a parallel between the structure of the normativity of meaning, and the structure of the normativity of reasons,

> to say that R is a reason for A is to say that one should do A because of R; and this requires two, a legislator to lay it down, and a citizen to obey (SN 137-8).

As Gert points out, Korsgaard's aim in drawing this parallel between the normativity of meaning and that of reason is to show that the publicity of reasons is analogous to the publicity of meaning. In this way, what the publicity of reasons is supposed to be like, will depend on what the publicity of meaning is.

We have seen what Gert thought the publicity of meaning consisted in: that the proper or improper use of the rules of language can be publicly observed; and that those rules can be taught. Correspondingly, the analogous publicity of reasons would involve the public observation of whether an action is the result of following a rational command (Gert 2002, p.313), and it would enable one to communicate one's reasons to others (Gibbard 1999, p.162).

If one thinks, as the authors endorsing the standard reading do, that the conclusion of this argument is that reasons are public; and one additionally thinks that the sense of the publicity of reasons Korsgaard seeks to establish is that their articulation does not contain an essential index to a particular person; then one would think that the idea that reasons are public in the sense that they can be observed and communicated is supposed to yield the idea that reasons are public in the sense that their articulation is not essentially indexed to a particular person.

If we read Korsgaard's argument that way, we would think that it fails in its aim, just as the authors endorsing the standard reading think it does. As Gert, in particular, reads the argument, reasons are public in the sense that third parties can check whether a given action is the product of following a given reason.[13] But that sense of publicity is unconnected to whether the specification of the given reason makes an essential reference to the agent or not (Gert 2002, p.313). Suppose we believe that you can check whether the

[13] A Kantian like Korsgaard, would be particularly averse to the claim that we can see whether an action – especially, another's action – has followed a reason. Whether an action has followed a reason depends on whether that reason was incorporated in the maxim behind the action. The maxim behind an action is not something that third persons can access easily (GW 407; SN 144).

reason why I am wearing my brown boots is that I wanted to look smart. The articulation of that reason would involve something like '...that I wanted to look smart', but that remains essentially indexed to me. So, even if we concede that you might be able to check whether a given action of mine is the product of following a given reason, that reason might perfectly well be agent-relative.

As Gibbard and Ridge point out, if by shareability we mean publicity (and by publicity we mean agent-neutrality), then the idea that the communicability of reasons entails their shareability seems to fly against a host of counterexamples – instances when you communicate your reasons to me, or a stranger does, and whilst I hear you and understand you, I don't thereby come to share those reasons (Ridge 2005b, p.70) – I might not even think that they should be reasons for you (Gibbard 1999, pp.162-163); nor does their articulation lose the fixed pronominal (Ridge 2005a, p.70).[14]

The authors endorsing the standard reading conclude that Korsgaard's argument trades on an ambiguity in the notion of publicity (Gert 2002, p.313). Reasons can be said to be public in the sense that they are communicable and (at least according to Gert) publicly verifiable. But that is a different type of publicity from agent-neutrality (Wallace 2009, pp.482-483). The argument from the publicity of reasons, therefore, fails to hit its target, according to this reading.

2.5.3. The permeability of consciousness

Korsgaard employs an additional argument to defend her thesis of the publicity of reasons. This argument appeals to the permeability of consciousness - the easiness with which our reasons intrude into each other's consciousness. Gert views the dialectical role of this argument as providing reinforcement to the argument from the publicity of language by arguing for the same conclusion. That conclusion, according to Gert, is that reasons are public in the sense of agent-neutral, and that that enables us to share our reasons.

[14] According to the reading of Korsgaard I favour, it is the case that the communicability of reasons entails their shareability. But, as we shall see, my reading of both the basis of that connection, and the way in which those reasons are shared, differs from the standard reading.

In her argument for the permeability of consciousness Korsgaard draws attention both to the ease with which we intrude into each other's consciousness, and to how we respond to those intrusions. By speaking to you, by calling your name, by asking you a question, or by making a request, I am interfering in your thoughts (SN 140).

This type of interference is different from the way in which events in our environment interfere in our thinking flow – the music next door, a leaf flying past the window. Unlike with interferences by those events, interferences by other agents are experienced as claims on us, as having a normative force. We typically respond to addresses by other agents by doing something simply because of the claim which those addresses place on us.

That claim might be intentionally or unintentionally placed. An example of the claim being intentionally put is if you ask me the time. There you are directly asking me to do something for you. You are directly placing a claim on me, and I will respond by telling you the time, or, if I can't do that, I will tell you that I can't. An example of a claim unintentionally put is where you announce that you are about to have the chocolate cake and I know that the cake is poisoned. There, I will feel a claim on me to alert you not to eat the cake. If I can't tell you, I will feel a need to justify it to myself and/or to you.[15] That is, I either comply with your request, or produce a reason why I don't.

Korsgaard's point is that for me to respond to your address in this way - either by doing what you say, or by giving you a reason why I don't do what you say - I must see you as having a normative status to me. This must be so, Korsgaard argues, because both those responses are responses which we issue to normative claims only. That is, our doing what is requested of us is not, or not solely, the output of calculated cooperation. If it was solely the output of calculated cooperation we would not feel compelled to do it (SN 139). And we feel a need to issue reasons only to normative claims (SN 140). So, the type of claim placed on us by another's address is a normative claim.

The idea that we give reasons to one another simply by interfering with each other's consciousness is crucial for Korsgaard's case for public reasons, and I will come back to it when I develop my own reading. Gert's own take is that the implication which Korsgaard wants us to draw from it is that if we give reasons to one another simply by interfering into

[15] The purported fact that I will feel a need to justify it to myself is on a par with the need to explain to you that I can't tell you the time. We can think of the first case as an instance of addressing a Strawsonian vicarious attitude (Strawson 1962).

each other's consciousness, then reasons are shareable, and thus public – i.e. agent-neutral (Gert 2002, p.314).

As Gert points out, that implication just doesn't follow. In order for your reasons to be reasons for me, more is needed than your attracting my attention to your reasons. For your reasons to be properly said to be reasons for me upon your communicating me them, they would have to meet two conditions. First, they would have to be properly articulable in agent-neutral terms; second, I will have to see them as having normative force for me (Gert 2002, p.314).

To be sure, these conditional circumstances *might* obtain. Here is one such case. We are both citizens of the same country, you explain to me the reasons why you are going to vote, and I conclude that those are reasons for me too. There we might say that your reason, which is an agent-neutral reason as it is not essentially indexed to you, but to anyone living in our country, becomes my reason too. But often, these conditions don't obtain, nor are they required to. You tell me your reasons for voting, but I live in a different country where those reasons don't apply, so your reasons don't become my reasons.

Gert's criticism of the argument from the permeability of consciousness, then, shows not so much that Korsgaard's analysis is wrong, as that the occasions described by her analysis are limited by pre-existing shared values between the agents involved. And this, of course, allows for the existence of agent-relative reasons – for reasons whose content is normative to you, but not for me. Korsgaard, then, has again failed to offer adequate support for her thesis that reasons are essentially public in the sense of agent-neutral.

In conclusion, the authors endorsing the standard reading agree that Korsgaard's arguments for the thesis that reasons are essentially agent-neutral fall short of their target. The argument from the publicity of language trades on ambiguities between 'public' in the sense of communicable, 'public' in the sense of shareable, and 'public' in the sense of agent-neutral. And the argument from the permeability of consciousness at most gives us an analysis of certain reasons which we happen to share, but does nothing to show that all reasons are shared in that way, nor that they ought to be shared in that way.

I think there is little doubt that according to the standard reading Korsgaard's thesis of the publicity of reasons is unsuccessful. However, I think that the standard reading misunderstands Korsgaard's conception of public reasons. That is, I believe that Korsgaard's

notion of the publicity of reasons is not equivalent to the notion of agent-neutral reasons. This misunderstanding of the conception of public reasons leads to a misinterpretation of Korsgaard's arguments and overall position. In the next chapter, I develop an alternative conception of the notion of public reasons, and a new reading of Korsgaard's discussion, based on that conception.

- 3 -

Korsgaard's public reasons as relations between agents

3.1.1. Introduction

In the previous chapter we saw that Korsgaard's discussion of the publicity of reasons does not come out well under what I have been calling 'the standard reading'. However, I believe that the construal of the notion of public reasons employed in the standard reading is mistaken, and that that leads to a misinterpretation of Korsgaard's arguments and overall position. In this chapter I develop an alternative conception of the notion of the publicity of reasons, and with it I a different reading of Korsgaard's discussion.

The standard reading takes Korsgaard's public reasons to be equivalent to agent-neutral reasons. Given that Korsgaard's sole explicit indication of what she has in mind by 'public reasons' says that it is '*roughly* what in contemporary jargon are called … "agent-neutral" reasons' (SN 133, n.3, my emphasis), it is not surprising that readers would have inferred that her public reasons are to be taken as agent-neutral reasons. But there are other considerations within the wider context of her discussion which suggest a different conception of public reasons.

Specifically, given that for my reasons to be reasons for you is for my reasons to be normative for you, it is right that we take heed of Korsgaard's account of normativity and of reasons. Once we are clear on her account of normativity and of reasons, we can see how those might 'spread' to others, and refocus our interpretation of Korsgaard's arguments - the arguments from the publicity of language, and from the permeability of consciousness - accordingly. In addition, we need to bear in mind the role which the concept of public reasons is supposed to play in establishing morality, viz., to show that valuing oneself 'implies, entails, or involves' valuing others (SN 132).

In Sections 3.3.1 to 3.4.3. below, I explore those contextual considerations in some detail. I will conclude that the conception of the publicity of reasons they support is the following: *reasons are public in the sense that what makes a reason a reason for one also makes it a reason for others.*[1] In this way, if I have a reason to ϕ, what makes it a reason to me also makes it a reason to you – not a reason for you to ϕ, but a reason for you to promote my ϕing. This is not to say that you will, or ought to, promote my ϕing - these reasons do not have to be overriding - but only that you are under normative pressure to do so (SN 140). I will say more about this later. From Section 3.5.1. to the end of the chapter, I will revisit Korsgaard's arguments from the publicity of language and for the publicity of reasons in light of this alternative conception of public reasons.

Under the reading I advance, the idea of the publicity of reasons does not imply that some reasons are public and others aren't, but rather that being public is a property of reasons *simpliciter*. If something is a reason, it has that property. This leads us to another aspect of the thesis of the publicity of reasons as I read it. The type of publicity in question is ontological, not epistemic. So, if something is a reason it has to be able to be a reason for others.[2]

Notice also that this does not yield an agent-neutral conception of reasons. As will become clear, A's reasons becoming reasons for B relies on a relationship between A and B. This relationship is captured in the articulation of the reasons in question by essential indexicals to the person whose reasons they are and/or have become. It is *my* reason that is a reason *for you* in virtue of a relationship in which we stand. So, instead of agent-neutral reasons (Cf. Gert 2002), public reasons are agent-relative reasons, only relative to everybody (providing some conditions, as we will see).

On my view, to understand Korsgaard's notion of the publicity of reasons, we have to understand her account of reasons. And in order to understand her account of reasons, we have to understand her account of normativity. So that is where I start.

---

[1] Wallace is an exception amongst the standard readers in noting that the publicity of reasons is supposed to consist in what makes them normative to one making them normative to others (Wallace 2009, p.471). However, Wallace goes on to attribute to Korsgaard the view that what gives reasons their normativity is the state of affairs which they describe (Wallace 2009, p.471). Viewed that way, Wallace is led to raise the same objections to Korsgaard's discussion as Gert did. However, as we shall see presently, that attribution goes contrary to Korsgaard's account of reasons as she develops it in Lecture 3 of SN.

[2] The epistemology of the thesis of public reasons is indirectly addressed by the discussion about the permeability of consciousness.

3.2.1. Preliminaries

Before launching into Korsgaard's account of normativity, a couple of preliminaries are in order. Firstly, my aim is not to provide an apologetic of Korsgaard's account of normativity or of reasons. Rather, my aim is to present a charitable exposition of her account of normativity and of reasons, in order that we might understand her conception of public reasons.

Secondly, presenting Korsgaard's account of reasons will inevitably take me over some of the same territory covered in the Chapter 1, especially the account of practical reasoning I outlined there. However, the account of practical reasoning I outlined there was a generic constructivist one. Korsgaard's own is sufficiently distinctive that *it* supports the account of public reasons she develops, whilst the generic constructivist account wouldn't, or at least not clearly. Since our purpose in this chapter is to gain an understanding of Korsgaard's conception of public reasons, we need to be clear about her account of reasons, even if this involves re-running some of the territory from Chapter 1.

3.3.1. Korsgaard's account of normativity[3]

In Chapter 1 I said that the constructivist's focus on normativity is directed to the agent's experience of normativity. Normativity is experienced by the agent as a kind of claim on the will - a distinctive 'pull' to do something. What marks normativity apart from other claims on the will, such as desires, is its origin: according to the constructivist, that is the agent herself, or agency (Sections 1.1.2. and 1.2.3.). That is what makes normativity different from other claims on the will - it is what makes it authoritative.

According Korsgaard, normativity arises out of the reflective structure of our minds. In being the kind of reflective creatures we are we can become aware of our mental states and activities - desires, perceptions, incentives, and so on. If these states or activities prompt us to do something, once we have become aware of them, we have to decide whether to do as they prompt us to do or not (SN 93). If what we are prompted to do is to believe something

[3] What follows is my reconstruction of Korsgaard's account of normativity. Some of the key steps are not much discussed by Korsgaard, so I have sought to draw sympathetic inferences from the steps she does discuss in detail, as well as from her overall picture. The sections of the following discussion which contain mostly my own reconstruction will be evident by the lack of references to Korsgaard's work.

S, then in questioning whether we should believe S we occupy the theoretical standpoint. If instead we are prompted to perform an action, then in questioning whether we should perform that action we occupy the practical standpoint.[4] This questioning marks a certain separation between the agent and the mental states or activities of which she has become aware. The agent sees the mental states or activities in question as different from her.

This separation exists within the reflective stance only. A scientist examining your brain might be able to identify the neural patterns involved in your having an impulse to do something, as well as your questioning whether to follow that impulse, and your final decision. From the point of view of the scientist, all of those activities are part of you, or of your brain. But this is not what it is like for you when you are in the reflective stance. From your reflective stance, there is a mental state or activity, which you are considering; and there is you, i.e., that which does doing the considering.

We can put it that when we become aware of an impulse to do something, we pose the question of what to do - the question of whether to follow that impulse or not. In posing the question of what to do, we manifest two aspects of ourselves: a commitment to producing an answer to that question, and a commitment to doing what the answer to that question says to do. These two commitments are what Korsgaard terms the thinking self, and the acting self.[5] Implicit in the thinking self - in the commitment to produce an answer to the question of what to do - is the presupposition that it is the agent that will produce that answer. That is, the answer to the question of what to do can't be settled by one's desires or impulses. Since the question is posed in reflection, the answer can only come from reflection (SN 93). The thinking self is the part, or aspect, of the agent that will bring forth that answer.

[4] By distinguishing between the theoretical and practical standpoints I mean neither to position myself on the debate of whether the two standpoints are as different from each other as it has been thought, nor to attribute any such positioning to Korsgaard. The main difference I wish to capture is that between thinking about what to do when I am not committed to doing whatever I conclude, as when I am conjuring up examples, thought experiments, of fiction; and thinking about what to do when I have to decide what to do in order to do something.

[5] Korsgaard does not characterise the thinking self and acting self as commitments of ourselves manifested in the question of what to do. She says of the thinking self and acting self that they are parts of us involved in the generation of normativity. That the job of the thinking self is to decide whether the agent is to do something or not (SN 107), and that the acting self concedes to the thinking self this 'right to government' (SN 104, 107). I believe that her overall position as I develop it here, suggests the characterisation of the thinking self and acting self I proffer here. I also think that this characterisation enhances the congruity of Korsgaard's position as a whole.

To produce an answer to the question of what to do, the thinking self scrutinises the idea of doing what the particular inclination in question prompts us to do. Let me explain. If I were to do as my inclination prompts me to do, I would do so under a given conception. Eg. 'I shall have the ice-cream because it would make me happy', 'I shall read the newspaper because I want to know what happened yesterday in parliament', 'I won't return your weapon because you have gone mad and may hurt someone'.[6] These conceptions of my prospective actions describe those actions as I, the agent, see them both in their empirical aspect, and in the way in which they are valuable for me. In the Kantian literature, they are known as my maxims.

Maxims are important for the identification and individuation of actions. For example, if I don't return your weapon because I want it for myself, that action would be different from the action where I don't return your weapon because I fear that you might hurt someone. The difference is in why I don't return your weapon. That difference - why I don't return your weapon - is not empirically manifested in the way that my not returning your weapon is. Instead, it features in my conception of my action - in my maxim.[7]

When I say that the thinking self scrutinises the idea of doing what the particular inclination in question prompts us to do, I mean that the thinking self scrutinises the maxim in question. This scrutiny results either in an endorsement of that maxim, or in an endorsement of the negation of that maxim (I explain what determines that endorsement in Sections 3.4.1. to 3.4.3.). Upon reflecting on whether to keep your weapon because you might hurt someone, I decide either to keep your weapon, or to not keep your weapon.

Given the nature of the acting and of the thinking selves, and the way in which they stand to each other, when the thinking self finds an answer to the question of what to do, it presents the agent with the opportunity to realise her commitment to do what the answer to the question of what to do says to do. That is, if the agent is committed to doing whatever the answer to the question of what to do is, and the answer to the question of what to do is to keep your weapon, the agent has the opportunity to realise her commitment to doing what the answer of what to do says to do, by accepting the decision

[6] This last example is Plato's, and used by Korsgaard (SN 108).

[7] Korsgaard explains the significance of maxims in a different way (SN 107-108). I use a different characterisation from Korsgaard's because I judge that it fits better with the exposition of Korsgaard's work I am developing.

to keeping your weapon. This makes the answer to the question of what to do normative for the agent. It makes it her reason for action (SN 97).

In Chapter 1 I said that the constructivist's focus on normativity was mostly on the agent's experience of normativity. The agent's experience of normativity is as a distinctive claim in the will to do something. This distinctiveness is important. If normativity is a claim in the will it has to be distinctive, otherwise it would be on a par with desires and desire-like states. I said in Chapter 1 that what gives normativity its distinctiveness over other claims on the will is that it comes from the agent. Korsgaard's account gives us a more detailed sense of the way in which the distinctiveness of normativity as a claim on the will comes from the agent.

The normative claim on my will to keep your weapon comes from myself in virtue of a relation in which I stand to myself. This is a relation in which I am authoritative to myself. It is manifested through my thinking self and acting self, which together form the question of what to do, and answer it. So, in posing the question of what to do - in occupying the practical standpoint - I pose myself in a normative relation to myself. This is the source of practical normativity, according to Korsgaard.

Let us recap Korsgaard's picture of normativity and of reasons so far. According to Korsgaard, the reflective nature of our consciousness is such that it forces us to separate ourselves from our mental states and activities. Where those mental states or activities prompt us to do something, we pose the question of whether to do as they say to do or not. In posing the question of what to do, we manifest a commitment to produce an answer to that question, and to do as that answer says - these commitments are our thinking self and acting self. The answer to the question of what to do consists in the reflective endorsement by the thinking self to do or not to do as the mental state or activity in question prompts us to do. The relation which the agent has to herself and which is expressed in through the thinking self and acting self makes the reflective endorsement of the thinking self normative - it makes it the agent's reason.

This relational structure of normativity and of the issuing reasons will be crucial for my interpretation of Korsgaard's notion of the publicity of reasons. But before moving on to that task, we need to bring to bear another aspect of Korsgaard's account of reasons. That is how the agent determines what is a reason - what is the process which the thinking self follows before it delivers its pronouncements. This will be especially relevant when it comes

to assessing how Korsgaard's attempt to establish morality fits within her overall account of normativity - my task in Chapters 4 and 5.

3.4.1. Practical identities

To see how the agent concludes the process of deliberation - how she decides whether to endorse a maxim or not - we need to say more about the beginning of the process. I have already emphasised in the above section, that when the agent occupies the reflective stance, for the agent it is as if she is distinct from her mental states and activities. As things are from the reflective standpoint, the agent is the one scrutinising the possibility of doing as her mental states or activities say to do. This way of conceiving of her situation demarcates the agent from her mental states or activities.

According to Korsgaard, another aspect of our reflective mind is that it forces us to have a conception of ourselves (SN 100). In this way, when I occupy the reflective stance and think of myself as different from my mental states and activities, I think of myself under a given self-conception. Equally, when I pose the question of what to do, what I'm asking for is what I - under a certain self-conception - am to do. These are self-conceptions are under which we 'value [ourselves] and find [our lives] worth living and [our] actions worth undertaking' (SN 101). Korsgaard calls them 'practical identities' (SN 101). This excludes descriptions which we self-attribute but that don't play a role in our normative landscape, eg. I have brown eyes, prefer the city to the countryside, and I was born in August. These descriptions are true of me, but they don't matter to me in the the way Korsgaard describes. In contrast, the following might be common practical identities: being a citizen of a given country, a vegetarian, a member of a certain profession, a supporter of a certain political ideology. These ways of seeing oneself do matter to people.[8]

If when adopting the practical standpoint I conceive of myself in one of these ways - under a practical identity - then in posing the question of what to do I will implicitly ask what I, conceived of in one of those ways, am to do. That poses a constraint on what would count as an answer to the question of what to do. The answer to the question of what to do will

[8] What self-attributed descriptions are or are not practical identities might vary depending on how socially causal those descriptions are. If a powerful ruler came along declaring that all brown-eyed people are inferior and that they should be exterminated, then my being brown-eyed would become a practical identity for me.

have to be expressive of myself under the given self-conception. That is, for me to reflectively endorse a given maxim, that maxim will have to be expressive of my practical identities. For example, if I am a vegetarian and I am wondering where to go for lunch, my being vegetarian will set constraints on what the answer to the question of where to go for lunch would be. A constraint in this particular case might be that I don't go to a stake house, for example. The answer to the question of what to do will be one expressive of my being a vegetarian. To put it in Korsgaard's pithy slogan, what reasons I have depend on who I think I am (SN 107).

3.4.2. The human identity

According to Korsgaard, then, our reasons spring from our practical identities. You have a reason to do whatever would preserve your practical identities. But we have multiple practical identities, so their requirements might conflict. You are a citizen of your country, and you are someone's daughter. Your country needs you to fight the fascists, and your mother needs you to look after her. You can't abide by both requirements. What are you to do? [9]

According to Korsgaard, practical identities are ordered in relation to how important they are for us. This ordering enables to resolve conflict between the requirements of different practical identities. When the conflicting requirements spring from practical identities of different value, we are required to uphold the requirement of the identity that is more important for us (SN 102). Korsgaard seems to think that not every practical identity is necessarily more or less important than all other identities. That is, she seems to think that a practical identity might be as important as another, or its importance might be incommensurate to the importance of other identities. In those cases the agent won't have access to that way of resolving conflict. As Korsgaard puts it, 'that's just one of the ways in

[9] This is an adaptation of Satre's well-known example of practical conflict. He didn't put the conflict in terms of practical identities, but the existentialist point he was touching on is similar to that implicitly alluded to by conflict between different practical identities.

which human life is hard' (SN 126).[10] (Not that deciding which of the conflicting practical identities is more important will always be an easy matter.)

The idea that we are required to do what our more important practical identities ask of us is troubling in another way. That is that, so long as there are no constraints as to what practical identities we might have, the requirement that we give precedence to our more important practical identities might yield unsavoury returns. If being a Nazi, or a white supremacist, are your most important practical identities, you will be *required* to do what those identities commit you to above anything else.[11] Korsgaard pre-empts this threat by arguing that there is one practical identity which we must all have and which is more important than any other practical identity: the human identity. This is how she argues for it.

Korsgaard's argument begins by considering the contingent nature of practical identities. The range of practical identities available to you is largely determined by the social settings into which you are born. As you grow up you might question the basis of those practical identities. That questioning might lead you to reject some and to acquire new ones. Or you might happily retain at least some of those identities for the rest of your life. The point is that practical identities might come and they might go. But *that* we have practical identities does not change. In other words, although we might change the self-conceptions under which we value ourselves, we must value ourselves under some self-conception or other.

It is necessary that we have practical identities because they supply us with reasons. And, according to Korsgaard's picture of the reflective mind, we need reasons to act. That is, once we pose the question of what to do, we need an answer to carry on - that answer is our reason. Without practical identities 'you will lose your grip on yourself as having any reason to do one thing rather than another - and with it, your grip on yourself as having any reason to live and act at all' (SN 121).

The requirement that we have practical identities at all lies behind the requirements of our particular practical identities (SN 121). If it didn't matter to me whether I had practical

[10] Korsgaard makes that comment with respect to whether moral reasons are always overriding of all other reasons. She thinks that they are not. Of course, we have not yet reached the issue of moral reasons. I take it that the comment is apt for the current discussion.

[11] As I mentioned in Chapter 1 this is a bullet some constructivists, especially Humean constructivists, are prepared to bite. Eg. Street, Lenman.

identities or not, it would not matter to me whether I failed to live up to my practical identities. But it does matter to us whether we live up to our practical identities or not. If I fail to live up to my being a pacifist, I will fail not only in my commitment to preserve that practical identity, but I will fail as a human being (SN 121). So, it does matter to us that we have practical identities. And that gives us a reason to have practical identities (SN 121).

But reasons come from practical identities. If our reason to have practical identities comes from our caring that we have practical identities, then the description that we need practical identities in order to live as agents, must be normative to us. That is, being the kind of creature which needs reasons to act must be a practical identity of ours (SN 121). This is the human identity. Since acting in accordance to reasons is the mark of agency, to have a human identity is to value oneself as an agent.[12]

The human identity is both necessary for agency and necessarily the most important identity of all. It is necessary for agency because without it we wouldn't have practical identities; without practical identities we wouldn't have reasons; and without reasons we can't be agents. And it is necessarily the most important identity of all because without it no other practical identity would matter.

If the human identity is both necessary and necessarily the most important identity of all, then Korsgaard's picture avoids the threat of relativism. Our practical identities must be expressive of our human identity.

Notice that through the human identity Korsgaard could answer Enoch's challenge. Enoch's most acute challenge to the constitutivist was to show that one is required to be an agent. Only then would the standards arising from agency be normative. Korsgaard has argued that we are required to value our humanity *if we are to act at all*. Her answer to Enoch is conditional, just as the general constructivist answer was (Section 1.3.5.). It is conditional on being an agent. So, as the conclusion we reached in Chapter 1, Enoch's shmagent, in not being an agent, would not be required to value his agency. That is both because he is not an agent, and because, in not being an agent, he is not subject to any requirements.

[12] The idea that Korsgaard has established that we must value our own humanity has been contested, amongst others, by Ridge (Ridge 2005b). Engaging with that discussion would divert me too much from my main task of explaining Korsgaard's conception of public reasons.

The human identity enables Korsgaard to avert relativism. One of the reasons why authors seek to avert relativism is that relativism threatens the very idea of morality. But whether the human identity is enough to institute morality will depend on what is required to maintain the integrity of the human identity. As we will see, Korsgaard's account of the publicity of reasons is intended to show that the integrity of the human identity requires that we treat others as ends.

But we are not done yet setting out Korsgaard's account of practical normativity. We still need to see how the agent reaches her conclusion - how she decides what maxims to reflectively endorse. The notion of practical identities gave us a formal answer to that question: since in posing the question of what to do the agent seeks to preserve her integrity under a given self-conception - a practical identity - the agent will reflectively endorse maxims which preserve the integrity of her practical identities. But it will be useful for us later to have a more substantive answer - an answer that tells us what it is for a maxim to preserve the agent's practical identities. This is Korsgaard's account of practical reason.

3.4.3. Practical reasoning[13]

In sections 3.4.1 to 3.4.2 above I introduced Korsgaard's notion of practical identities as self-conceptions which furnish one's sense of self. In this way, when posing the question of what to do, we ask what to do within the normative parameters set by one's practical identities. The answer to that question will have to be one which stays within those parameters. That is, the maxim proposed as an answer will have to preserve the integrity of the agent as per the descriptions under which she values herself.

Part of those parameters is that the agent herself produces the answer to the question of what to do. That is so because practical identities are constituted by commitments of the agent. As such they can only be preserved through upholding those commitments. And those commitments can only be upheld by the agent herself.

To see why, think about what it would take for you to continue being a vegetarian. It would not do if the decisions to choose vegetarian meals were made by someone else on your

[13] This section is a reconstruction of Korsgaard's discussion in SN 107-113.

behalf - maybe your partner is herself vegetarian and won't allow any other food in the house, or an evil scientist has manipulated your brain in such a way that any thoughts about food automatically and irresistibly lead you to reach for the vegetables. None of these ways of arriving at your 'decision' about what to eat would count as expressions of yourself as a vegetarian, nor would they contribute to the integrity of yourself as a vegetarian. To be sure, they would have the same consequences, namely, that you don't eat animals. But that is not all it takes for someone to be, and to stay, a vegetarian. So part of the requirements of practical identities is that the decision of what to do is made by the agent herself.

So the agent's answer to the question of what to do will have to preserve the integrity of the agent's practical identities, and, for it to do that the answer will have had to be produced by the agent herself. The agent will have to decide whether the maxim in question is expressive of her practical identities. How the agent decides whether a given maxim is expressive of her practical identities depends on what it is for a maxim to preserve the integrity of the agent's practical identities. Let's look at that now.

We saw in Section 3.4.1. that practical identities are constituted by commitments of the agent - their descriptive content is normative for the agent in that she is committed to uphold it. For example, if the descriptive content of being a vegetarian is that one chooses not to eat animals, the agent who has that as a practical identity is committed to choosing not to eat animals. We might view those descriptions to which the agent is committed as principles. In this way, choosing not to eat animals is a principle of being a vegetarian; discouraging the use of violence is a principle of being a pacifist; engaging in the social and political life of that country is a principle of being a citizen of a given country.

For a maxim to preserve the integrity of the agent's practical identities is for the agent to be able to will that she adopt that maxim as principle without violating the principles constitutive of her practical identities. Principles are universal. So for the agent to be able to will that she adopt that maxim as a principle without violating the principles constitutive of her practical identities, is for the agent to be able to will the universalization of the given maxim without violating the principles constitutive of her practical identities.

For example, suppose that the action which the agent is considering is captured in the maxim 'joining the army in order to protect my country'. For this maxim to be a principle is for it to say what one is to do whenever one finds oneself in the current circumstances. That is what it is for it to be universal. So, for me to be able to will the universalization of that

principle is for me to be able to will that I commit myself to joining the army in order to protect my country whenever I face the current circumstances.

The agent endorses whatever the outcome of the test of universalization is. If I cannot will the universalization of the given maxim, I shall endorse not acting on that maxim; if I can will its universalisation, I shall endorse acting on that maxim. If I am a pacifist, I cannot will that maxim as my principle because doing so would violate the principles constitutive of my practical identity. In this case I would endorse *not* joining the army in order to protect my country. If am not a pacifist I might be able to will the universalization of that maxim, and in that case I might endorse joining the army in order to defend my country.[14]

This, then, is how the answer to the question of what to do is produced. The process which the agent follows to reach that answer is practical reason, and the test of universalization is the principle of reason.

To test one's maxims for universalization is to apply the categorical imperative test, and to act on whether one can will one's maxims to be universal is to follow the categorical imperative: 'act only in accordance with that maxim through which you can at the same time will that it become a universal law.' (GW 4:421)

But Korsgaard's construal of the categorical imperative differs from Kant's own. According to Kant, the categorical imperative is the principle of reason *and* the moral law. Kant thought that the categorical imperative is the moral law because he thought that the scope of universalization of the categorical imperative comprised all agents. That is, Kant thought that when applying the categorical imperative the agent tests whether she can will that *everyone* makes the given maxim their principle. In contrast, according to Korsgaard, the scope of application of the categorical imperative is the agent herself. The agent tests whether she can will that *she* makes the maxim in question her principle.

Korsgaard departs from Kant on the scope of application of the categorical imperative because that scope comprises the sources of normativity for the agent. Kant, like Korsgaard, thought that the agent is normative to herself - that she necessarily sees herself as an end. But, unlike Korsgaard, Kant thought that in showing that the categorical imperative is the

[14] I say I *might* be able to will the universalization of that maxim if I'm not a pacifist because other more important practical identities of mine might make different requirements, for example, I might need to take care of my mother.

principle of reason he had shown that all other agents are sources of normativity for the agent. Korsgaard agrees that all other agents are sources of normativity for the agent. But she thinks that that is not established by showing that the categorical imperative is the principle of reason. An additional step is needed. Her account of the publicity of reasons is supposed to supply that additional step.

3.4.4. Summary of discussion so far

After this lengthy presentation of Korsgaard's account of normativity let us situate her account with respect to morality. The moral challenge to constructivism is to show that the agent is to treat others, as well as oneself, as ends. That is a challenge because, in order to be required to *treat* anyone as an end, one has to *see* them, to *value* them, as an end. Yet it looks like according to constructivism one cannot value others as ends. That is because to value another as an end is to regard them as setting ends to oneself regardless of one's interests. However, according to constructivism one is required to preserve one's most important valuings. The most important valuing of all is one's valuing of oneself. So it looks like the agent cannot see others as ends. Her regard for others must always be subordinated to her valuing of herself.

That was also Kant's problem with his Formula of Humanity. The Formula of Humanity also instructs us to treat others and ourselves as ends, but critics have argued that Kant's argument for the Formula of Humanity only establishes (at most) that one is to treat oneself as ends. And it looks like there is no route available for Kant to establish that one must treat others as ends.

Korsgaard's dispute with Kant's conception of the categorical imperative takes us to the same point. As I have just explained, Kant thought that in establishing that the categorical imperative is the principle of reason he had established that it is the moral law. He thought that because he believed that the principle of reason had all agents as sources of normativity for the individual agent. But Korsgaard has added to the discussion by pointing out that the categorical imperative as the principle of reason has only the agent herself as her source of normativity. Here too we find ourselves valuing only ourselves and without a clear way to show that we are to value (hence treat) others as ends.

In Section 2.4.1. I anticipated that Korsgaard's attempt to establish a constructivist morality would, if successful, vindicate Kant's Formula of Humanity. Her strategy, I said, is to argue that valuing oneself 'somehow implies, entails, or involves' valuing others (SN 132). Now we can see that her strategy would also establish the moral law. The moral law is the categorical imperative applied to all agents. Since the scope of application of the categorical imperative comprises the sources of normativity for the agent, if Korsgaard shows that the agent values all other agents as ends - i.e. if the agent regards all other agents as sources of normativity for her - her application of the categorical imperative will comprise all other agents. It all rests on Korsgaard's ability to show that valuing others follows from valuing oneself.

In this chapter I have presented Korsgaard's arguments for the view that one must value oneself. Our reflective mind is such that we need reasons in order to act. In order to have reasons we need to value ourselves under certain conceptions - we need to have practical identities. And in order to have practical identities we need to value ourselves as agents - as ends in ourselves. Through the thesis of the publicity of reasons Korsgaard is supposed to show that valuing oneself leads to valuing others.

In Sections 2.4.1 to 2.5.3., I presented what I called the standard reading of Korsgaard's discussion of the publicity of reasons. According to the standard reading reasons are public in the sense of agent-neutral. We saw that under the standard reading Korsgaard's arguments are riddled with equivocations and overstretched conclusions. I proposed to develop an alternative reading by tending to Korsgaard's wider discussion, specifically her account of normativity and of reasons.

I said that my reading of Korsgaard's discussion has at its core the idea that reasons are public in the sense that what makes a reason a reason for me also makes it a reason for you. This is to say that what makes a reason normative for me also makes it normative for you. I shall next argue that that construal is supported by Korsgaard's account of normativity and of reasons. The support available is not unequivocal. But I hope it is sufficiently plausible to make it seem a worthy alternative to the standard reading.

3.5.1. Public reasons redux

According to Korsgaard, my reasons are my answers to the question of what to do. In other words, my reasons are instructions which I give to myself in virtue of a relation in which I stand to myself - a relation involving my thinking self and my acting self. Those instructions are normative, i.e. they are reasons, because of yet another relation in which I stand to myself and which underpins my thinking self and acting selves. That is the relation in which I regard myself as normative to myself, as an end in myself. That is my human identity. The pronouncements, or reflective endorsements, of my thinking self are normative for me because I recognise them as pronouncements of the thinking self - as issuing from my commitment to producing an answer to the question of what to do. This is a thumbnail sketch of Korsgaard's account of reasons and their normativity.

I have already indicated that, save for the footnote where Korsgaard says that what she is 'calling "private" and "public" reasons are roughly what in contemporary jargon are called "agent-relative" and "agent-neutral" reasons' (SN 133, n3), Korsgaard does not give us a clear and explicit characterisation of the notion of public reasons. However, she does make some statements which, when read in light of her account of reasons, *and* in light of the role she explicitly assigns to the publicity of reasons, do point at a specific construal of the notion of the publicity of reasons.

I list those statements so I can then keep track of how they each fit in the construal I develop:

(i) If reasons are private their normative force applies only to the agent who has produced those reasons (SN 133).

(ii) If reasons are public their 'normative force may be shared with others' (SN 136).

(iii) For me to share into your reasons involves for me to 'make *your* humanity normative for me' (SN 134, original emphasis).

In light of Korsgaard's account of normativity and of reasons, I propose the following interpretation of those statements.

- If under a private conception of reasons the normative force of reasons applies only to the agent in whom those reasons originate (i), and the normative force of reasons

comes from the agent's relation to herself as a source of normativity (the human identity), then under a private conception of reasons, you can be normative only for yourself.

- If a private conception of reasons contrasts with a public conception of reasons, and under a public conception of reasons their normative force may be shared with others (ii), then under a public conception of reasons you can be normative for others, as well as for yourself. That is, you can be an end for others, as well as for yourself.

- If for me to share into your reasons involves for me to make your humanity normative for me (iii), then for me to share into your reasons involves for me to see you as an end for me.

All of Korsgaard's statements listed above, when read in light of her account of reasons and of normativity, yield the idea that the publicity of reasons is a matter of agents seeing each other as ends for each other. I see you as end for me, and in virtue of that, your reasons are reasons for me. Just as it is in virtue of seeing myself as an end for me that my reasons are reasons for me.

But showing that we must value others as ends is precisely what the constructivist has to show to establish morality (Section 2.2.1.). And it is the job which Korsgaard explicitly assigns to the notion of the publicity of reasons: to show that we are required to value others as ends. [15]

So, both an interpretation of Korsgaard's comments about the publicity of reasons in light of her account of normativity and of reasons, and her explicitly stated role for the notion of the publicity of reasons, converge on the same idea: reasons are public in the sense that their normative source is recognised by all agents. This is the sense in which I declared earlier (Section 3.1.1.) that reasons are public in the sense that what makes a reason a reason for one also makes it a reason for others. What makes my reason a reason for me is my self-

[15] Specifically, as I have stated repeatedly, the publicity of reasons is supposed to show that we are required to value others as ends by showing that valuing oneself involves valuing others. The question of how the publicity of reasons helps establish that connection will be examined in Chapter 4. In the current chapter, my aim is simply to articulate and defend a reading of the notion of the publicity of reasons.

regard as an end. What makes your reasons reasons for me is my regard for you as an end - my seeing you as an end.

I shall work with that construal of the notion of the publicity of reasons, and revisit Korsgaard's arguments from the publicity of language and for the permeability of consciousness in its light. My hope is that at the end of this chapter the construal of the notion of public reasons I propose is shown to be more profitable than the construal employed by the standard reading. I hope that that is so, even though in Chapter 5 I shall argue that it ultimately fails in its purpose to establish morality.

Before proceeding to revisit Korsgaard's arguments for the publicity of reasons, there are two caveats I should make. The first is that, if the construal of public reasons I employ is correct, Korsgaard's arguments from the publicity of reasons and for the permeability of consciousness won't be aimed to providing deductive support for the thesis of the publicity of reasons. Rather, they will be meant to show that that conception of reasons allows us to make more sense of our interactions, and that Korsgaard's account is amenable to such a conception of reasons. We will appreciate this more fully after we have gone through the arguments.

The second caveat is that the construal of the publicity of reasons I employ is both ambitious and unorthodox and, as is common in such cases, there would be much refining to be done for it to provide a comprehensive account of the way in which we are supposedly normative to one another. However, my main interest in the notion of the publicity of reasons won't concern the missing details, but its broad structure which is supposed to accommodate the very idea that we might be normative to one another. To pursue my interest, the incomplete account of the notion of publicity of reasons available to us will suffice.

3.6.1. Argument from the publicity of language

According to my reading, Korsgaard's argument from the publicity of language is designed to show that her account of reasons *allows* for the normativity of reasons to apply to other people. That is, she aims to show that on her account of reasons my reasons *can* become reasons for you. She tries to do that by drawing attention to the way in which her analysis

of reasons maps onto Wittgenstein's analysis of language. In that way, if we accept that Wittgenstein's analysis of language leads to the view that languages are public, we will be pressed to accept that Korsgaard's analysis of reasons leads to the view that reasons are public.

Wittgenstein was concerned with the question of what determines whether an utterance conforms to a rule. Broadly speaking, a rule says what you ought to do, and to conform to that rule is to do what that rule says to do because that is what the rule says to do. That just follows from the fact that rules are normative. So, conforming to a rule is a relational property. It is a matter of whether something A does/is what something B says for A to do/be.

Wittgenstein pointed out that an agent's use of an utterance *alone* could not in itself determine whether the utterance conforms to a rule. The agent's use of the utterance alone can only ever be what the agent *does*, and not what the agent *ought* to do. If that is the case, there is no question of whether the agent does what she ought to do. That is because - as far as the agent's use of the utterance along goes - there is nothing that she ought to do. In other words, there is nothing for the agent's use of the utterance to conform to. This shows that the agent's use of the utterance alone does not allow room for normativity.

Wittgenstein concluded that for there to be something that determines whether an utterance conforms to a rule, the rule and utterance have to originate in different places. Hence, meaning must be public - it must involve more than one party.

Korsgaard's formulates Wittgenstein's insight thus,

> to say that X means Y is to say that one ought to take X for Y; and this requires two, a legislator to lay it down that one must take X for Y, and a citizen to obey (SN 137).

This formulation is supposed to highlight that what accounts for the publicity of meaning is the relation between legislator and citizen at the heart of the structure of meaning. But that relation is just the relation that Korsgaard argues is constitutive of normativity (SN 137-138). On her account the role of legislator is played by the thinking self, and the role of citizen by the acting self.

Indeed, as Korsgaard points out, that formulation of meaning maps neatly onto Korsgaard's formulation of reasons:

> to say that R is a reason for A is to say that one should do A because of R; and this requires two, a legislator to lay it down, and a citizen to obey (SN 137-8).

I propose that the way to interpret Korsgaard's use of Wittgenstein's argument is this: if the relational nature of the structure of meaning is what accounts for the publicity of meaning, then since the structure of reasons displays the same relational nature, it follows that reasons must be public too.

We could put it in slightly different terms: if meaning is public because it is a normative phenomenon, then since reasons too are a normative phenomenon, they too must be public. Either way, the idea Korsgaard tries to support is the same: if another can play the role of legislator for me in the case of meaning, they can also play that role in the case of reasons.

As I intimated in my first caveat above (page 72), according to my reading, Korsgaard's use of Wittgenstein's argument for the publicity of meaning is not supposed to provide a deductive argument for the publicity of reasons. Rather, through it she seeks to highlight that her account has the theoretical capacity to accommodate the idea that my reasons might be reasons for others. What makes that possible is that, according to Korsgaard, reasons are generated through the interaction of two different (though related) aspects of ourselves which all agents have in virtue of being agents - viz. the thinking self and the acting self.

To put it another way, Korsgaard's aim is to persuade us that there is nothing on her account of reasons that prevents the possibility that someone's reasons become reasons for another. She argues that reasons arise from the interaction between thinking self and acting self. But nothing on her account determines that the thinking self and the acting self in question must belong to the same person. This idea is given more buoyancy by the argument for the permeability of consciousness, as we shall see in the next section. But I first want to bring this idea and why it matters, further into relief by contrasting it with a private conception of reasons.

As we saw earlier (Section 2.4.1.), according to a private conception of reasons, reasons are not formed through a relation which the agent has for herself; rather they are constituted by single mental state: a desire might be a reason, pain might be a reason, pleasure might be a reason. Mental states can only belong to the person whose mental states they are. Your desire can be only yours, the same as your pain and as your pleasure. If those mental states are reasons, it follows that reasons can never be shared. I can never share in your reason

because I can never share in your desire, or in your pain, or in your pleasure. Your desires, pain, and pleasure might *cause* desires, pain, or pleasure in me. That is, they might give me reasons. But just as the desires, pain, or pleasure which your mental states might cause in me will remain ontologically private, so will the reasons your mental states give me, for the mental states your reasons give me *are* my reasons.

In this picture, there is *no way* in which reasons can be shared. There is an insurmountable ontological obstacle for the sharing of reasons in this picture. That is that what is constitutive of reasons, namely, specific single mental states, can belong only to the individual.

In contrast, if reasons are relations, as on Korsgaard's account, and the parties involved in those relations (thinking self and acting self) obtain in all agents, then, *at least at that level of description*, the relations constitutive of reasons *can* obtain between different agents. As I said in my second caveat above (on page 72), much more would need to be said to complete an account of interpersonal normativity. Some more is still to be said through the argument for the permeability of consciousness. But through Wittgenstein's argument, all that Korsgaard aims to show is that the path is open for her account of reasons to develop into an account of interpersonal normativity, at a point where it isn't open for the rival account of reasons. We already accept that a normative phenomenon (viz. meaning) might involve different persons. Reasons, as a normative phenomenon too, can also involve different persons.

3.7.1. Argument for the permeability of consciousness

According to Korsgaard, then, reasons are public in that they are shareable. What makes them shareable is their relational structure. That relational structure is formed by a relation between the thinking self and the acting self. This is a relation where the acting self recognises the authority of the thinking self. The thinking self is the agent's commitment to producing the answer to the question of what to do. It is authoritative for the agent because it is an expression of the agent's valuing of herself as an agent - it is an expression of the agent's human identity.

The argument from the publicity of language provides an account of how it is possible that we are normative to one another. But to say that our reasons *can* be reasons for others is not to say that *are*, much less that they *ought to be*. I will deal with the question of whether they ought to be in Chapter 4 and, under different considerations in Chapter 5. Here, I look at how Korsgaard attempts to show that they are.

Korsgaard's efforts to show that reasons are shared consist in showing that we can give a more suitable interpretation of interpersonal relations under her conception of reasons than under the alternative. In effect, Korsgaard supports the idea that reasons are shared by putting the onus of proof on the sceptical camp.

Once again, my presentation here is a reconstruction of Korsgaard's discussion. I begin by considering what is involved, conceptually, in sharing each other's reasons, given Korsgaard's account. I then explain how Korsgaard applies the resultant model to the interpretation of interpersonal relations.

For your reasons to become my reasons - for me to share in your reasons - is for me to recognise the authority of your thinking self. It is important to get this point clear. If your reason becomes my reason, that is not because I am persuaded that your reason is correct. If that were how your reason comes to become my reason it would be a case of my recognising the authority *not* of your thinking self, but of *my* own thinking self: *I* decide that your reason is a good reason, so I adopt it. From my practical standpoint, this would not be a case of your reason becoming my reason, because *as far as I am concerned*, what we are calling *your reason* would not be a reason until I decide that it is a reason. Rather, for your reason to become my reason is for it to become my reason in virtue of my recognising it as a reason. In recognising your reason as a reason I recognise you as a reason-giver, as a source of normativity, as an end.

There are two main issues arising from this picture. The first is that this picture implies that for me to see you as an end, is for me to see you as an end *for myself*. The second is the question of how to account for the idea that your reasons become my reasons whilst I think that they are wrong. I take them in turn.

The first issue is that according to this account to see another as a source of normativity, as an end, is to see them as a source of normativity, as an end, *for oneself*. However, this is not a problematic or illegitimate implication. Rather, it follows from the phenomenological

aspect of normativity. As we saw in Section 1.1.2., normativity is experienced by the agent as a distinctive claim on the will. Given that, I could not tell that something/someone A is normative unless I experienced that A places a claim on my will. I might have the *concept* that A is normative, but I would not be able to tell whether that concept was veridical unless I could tell whether A is normative. And I can tell that only if I experience that A is normative.[16] So, the implication in the picture above that to see another as an end is to see them as an end for you is a legitimate one.[17]

The second issue arising from that picture is that, if it is the case that your reasons become my reasons *not* in virtue of my deciding that they *are* reasons (or good reasons), but in virtue of my recognising your authority, what about my judgements about the quality of your reasons, and/or about how they fit with my own reasons? It would be implausible to deny that we have judgements about others' reasons, and that those judgments influence how we respond to those reasons. To deny that would fly in the face of plenty of our experiences in interpersonal relations. A member of a pro-hunting charity asks me for a donation, but I think that she is wrong because I believe that hunting is wrong. Or, you call me but I'm running for an appointment, so I can't come to you (SN 140). Yet, if we say that our judgements of others' reasons influence how we respond to those reasons it looks like we are saying that whether others' reasons become our reasons is a matter of whether we judge that they are good reasons. But, as I explained above (page 74), that would not be a case of your reasons becoming my reasons in the sense of my responding to your normative status. It seems that the picture of what is involved in sharing into each other's reasons faces a dilemma.

Korsgaard's way out of this dilemma consists in drawing some nuances on how one might respond to another's normative status. We can infer from Korsgaard's discussion that she believes that there are two broad ways in which we might respond to another's normative status. Doing as the other's reasons say to do (because that's what they say to do) is but one

[16] One of the details missing from the account of the publicity of reasons presented here is the issue of misperception of another's normative status, or lack of it. This is not a problem here, however, as it does not affect my main object.

[17] This might raise the question of how to account for the notion that I recognise that someone's reasons, A's, are reasons for someone else, B, but not for C, or for me. I think that phenomenon can be accommodated by Korsgaard's account without much difficulty. Developing the required argument would take us too much out of the main purpose of this dissertation. But the argument would go along the lines that there are certain reasons which you want to be normative to some agents and not to others. *That* qualification is part of your reasons. Eg. you want chocolates, but you want your darling, not me, to get them for you. My not attempting to get you the chocolates follows from my understanding that you don't want me to get you them.

of those ways. The other is by *producing a reason* why I am not doing as your reasons say to do. I can explain to you that I'm busy, or that I think that your reasons are in error. According to Korsgaard, to offer you a reason when I can't do what your reasons say to do indicates that I am responding to your reasons as your reasons, that I am treating you as an end (SN 140).

Korsgaard does not expand on that point. But I think that the idea she is alluding to is the following. Only agents can appreciate reasons. And agents are sources of normativity. If I offer you a reason I am therefore treating you as a source of normativity. If I didn't think of you as a source of normativity, I would not think it appropriate to give you reasons, any more than I think it appropriate to explain to my tulips why I am moving them away from the windowsill. So, according to Korsgaard, issuing a reason why I don't do as your reasons say to do shows that I remain bound by them – more precisely, that I remain bound by the source of authority behind them, namely, you.

We might think that this answer begs the question. The question was whether I treat you as an end – whether your reasons are my reasons – when I judge that they are wrong, or that I can't do as they say to do. The answer I proposed is that, since in those cases I offer you a reason, I am treating you as an end. But why think that what I offer you is a reason? Why not think that I am simply saying things that I think will keep you in good terms with me? To think that what I am offering is a reason is already to assume that I see you as an end. But that is what we are trying to establish.

Korsgaard anticipates this challenge (SN 141). Her answer to it comes from her general argument that her account offers a richer and more recognisable analysis of our interpersonal interactions than the alternative account. We will see that when I lay out her analysis momentarily. Of course, even if Korsgaard's analysis persuades us that we do see each other as ends, we can still accept the possibility that we might treat others as means only. I might offer you a mock reason – something that looks like a reason but that is not because through it I am treating you instrumentally only. But accepting this does not involve ceding way to the challenge above. The challenge above, if it is to be a serious challenge, is not that we might treat each other solely as means; rather it is that treating each other as means is *all* that is available to us.

One difference that follows from that contrast is that to *treat* you as a means when I *see* you as an end would be *wrong*; but to treat you as a means where that is all that is available to

me, is not to do anything wrong. The challenge above raises the possibility of the latter scenario; Korsgaard's answer insists on the first.

Having sketched some aspects of the concept of sharing into each other's reasons, let's see how it helps us analyse actual interactions, in contrast with the analysis provided by the alternative construal of reasons as private reasons.

Suppose you and I are trying to decide where to go for coffee.[18] You suggest café Plató, but I respond that it is too out of my way. I suggest café Full Circle, but you respond that the cakes there are not very good, and you quite fancy a nice piece of cake. You suggest café Quarter Mile, and that is good both for me and for you. So we agree, and go to café Quarter Mile.[19]

If we think of reasons as private, in the sense that their normativity stays within the agent, we will analyse that exchange as follows. Each of us has a separate goal: your goal is to have coffee with me, and my goal is to have coffee with you. Your goal involves me, and mine involves you. But since in this picture goals will be (very broadly) types of desires, they are distinctively yours and mine respectively. Given our separate goals, we each go about what we think is the best way to achieve our separate aims of having coffee with each other. We might see each other's objections as practical obstacles to overcome, and we judge that the easiest way to overcome them is by suggesting alternative places to meet.

There are certain conditionals associated with this picture. If I thought that the easiest way to overcome your objection was by deceiving you into thinking that the cakes in café Full Circle are good enough, I would do that. If you thought that the easiest way to overcome my objection was by bullying me into thinking that my objection is not a good one, you would do that.

Our negotiations about where to go might be of a nicer nature. We might be fond of each other. In that case, our objections cause desires on each other which we seek to fulfil. Your objection to having a substandard cake stirs in me a desire that you have a good cake. With that end, I am willing to go wherever you think cakes are good, so long as doing so does not

[18] This example, and the ensuing discussion is a reconstruction of Korsgaard's discussion in SN pp.140-141.

[19] Korsgaard's example on which mine is modelled presents a conversation between a student and a teacher (SN 141). I don't use Korsgaard's own example because I believe that given the strict roles of the protagonists - a student and a teacher - it invites questions which, although important and interesting, make the task of conveying an elementary notion of the publicity of reasons 'in action' more cumbersome.

clash with my desire to not go too much out of my way. Your awareness of my desire to not go too much out of my way for our coffee, will cause you to have a desire that I don't go too much out of my way. You are prepared to ensure that I don't go too much out of my way, up to the point where doing so would clash with your desire to have a nice cake.

The point is that both in the nice and in the less nice interpretations supplied by the private conception of reasons we each are trying to satisfy whatever desire (broadly construed) is more commanding at the time. Whether that is expediency in settling where to meet, or the satisfaction of seeing a friend get what she wants (so long as it is at not great cost to ourselves). Under a conception of private reasons that is *all* that is open to us. We can treat each other only as means to our ends.

In contrast, if we think of reasons as public, the analysis we will provide is the following. You and I have a shared aim: to have coffee together. We can say that it is an aim we share because an aim is a kind of decision: we have decided to have coffee together. As such, it has a relational structure that allows for it to involve two persons. Our exchange about where to have coffee is one in which your objection becomes a reason for me, and my objection becomes a reason for you. In this analysis, it is not the case that we come up with alternative suggestions of where to go for coffee because we judge that that is the least costly way of attaining our aims. Rather, we come up with alternative suggestions because our separate objections have the status of ends for each other.

As in the private reasons interpretation, there are certain conditionals associated with this interpretation. If I thought I could easily bully you into having the less than excellent cake, I still wouldn't do it. If you thought you could manipulate me into displacing myself further than I would like, you still wouldn't do it. I wouldn't bully you, and you wouldn't manipulate me because doing so would be to treat each other as means, and that is not how we see each other. We see each other as ends. That is the way in which our consciousness is permeable to each other's normative status.

The idea behind contrasting the analyses provided by a private and by a public conception of reasons is to show that the analysis provided by a public conception of reasons is more recognisably true than that provided by a private conception of reasons. If the analysis provided by the publicity of reasons is thought to be superior to the alternative, then we have good reason to believe that as a general matter of course, we do share into each other's reasons. If our intuitions about which analysis is superior are not settled on the basis on this

example alone, we can look back to the example I used in Section 2.2.1. above, to illustrate the challenge of morality to constructivism.

The private reasons analysis of our discussion about a coffee place is similar to the analysis of my interaction with the ticket seller where I treat her solely as a means (Section 2.2.1.). That analysis of my interaction with the ticket seller was used to push the idea that, if that analysis is all that the constructivist could provide, the constructivist falls short both of accounting for morality *and* of accounting for my agential role as I experience it. My agential role, as I experience it, is that I treat the ticket seller in a certain way despite how that impacts the attainment of my ends. That is, my conception of my actions regarding the ticket seller is that I see her and treat her as an end. And that is what morality requires too: that we treat others as ends. Both my conception of my action, and the requirement of morality, corresponds neatly to the public reasons analysis of our discussion about where to go for coffee.

If in Section 2.2.1. we thought that constructivism was in trouble because it looked like it could only account for agents treating each other as means, and not as ends, we should conclude that the analysis of our interaction about where to go for coffee provided by the public construal of reasons is superior to that provided by the private conception of reasons. It is superior because it captures both my conception of my action, and the requirement of morality - namely, that I am treating you as an end.

As noticed above if the account of the publicity of reasons is correct, it can still be the case that we treat each other solely as means. But there are two related notable differences between treating each other solely as means when the construal of reasons as public is true, and treating each other solely as means when the construal of reasons as private is true. The first is that, if the construal of reasons as private is true, treating each other solely as means is the only way of treating each other available to us. Under a public conception of reasons, by contrast, we *see* each other as ends, and so we can *treat* each other as ends. So, unlike with private reasons, with public reasons we can treat each other as ends.

The second difference issues from the first. If the construal of reasons as public is true, then, as I said, we see each other as ends. Given what is involved in valuing something (Section 1.1.2. and 1.1.3.), and what it is for something to be an end (Section 2.2.1.), if we see each other as ends, we are required to treat each other as ends. Therefore, if we see each other as

ends, and not treat each other as ends, we are doing something wrong. Therefore, the public conception of reasons allows us to say that it is wrong to treat others solely as means.

If, in contrast, the privacy of reasons is true, since the only way in which we can treat each other is as means, there is nothing wrong with treating others as means per se. The only criticism available to how we treat each other concerns whether the way in which we elect to treat each other fits our preferences best.

This invites us to look at Korsgaard's account and arguments for the publicity of reasons from the following angle. Treating each other as ends is something which we do as matter of course, and it is what morality requires. The question is how to account for that phenomenon. A private account of reasons does not allow for the idea that we treat each other as ends. It only allows for the idea that we treat each other instrumentally. What would allow for the idea of treating each other as ends is a public conception of reasons - a conception of reasons in which my reason becomes your reason just because it is my reason (argument for the permeability of consciousness). But how can we account for public reasons? We can account for public reasons through the relational nature of normativity (argument from the publicity of language). The relational nature of normativity is founded on the reflective nature of the agential mind. So all agents can participate in the sharing of reasons.

This concludes my exposition of how the argument for the permeability of consciousness is supposed to show that, as a matter of course, reasons are shared in the way predicted by the thesis of the publicity of reasons.

3.8.1. Summary of Chapters 2 and 3

My task in these two last chapters has been to begin exploring whether the constructivist can account for morality. In Chapter 2 I argued that constructivism *seems* to not have the conceptual capacity to account for morality because it seems that it cannot accommodate the idea of treating others as ends. I then argued that Korsgaard's account of the publicity of reasons is an attempt to show that constructivism can accommodate that idea. I presented the standard reading of Korsgaard's discussion. According to that reading, Korsgaard seeks to establish that all reasons are agent-neutral, but her arguments fall short of that target.

In the current chapter I have articulated and defended a reading of Korsgaard's notion of public reasons according to which reasons are inherently shareable. All that has to be in place for reasons to be shared is for us to stand in a given relation to each other - a relation in which we see each other as ends.

Korsgaard's account of the publicity of reasons is supposed to show that valuing oneself involves valuing others (SN 132). In the next chapter I examine how the conception of the publicity of reasons I attribute to Korsgaard is supposed to establish that connection.

– 4 –

Public reasons and morality

4.1.1. Introduction

Morality requires that we *treat* others as ends. We can *treat* others as ends, only if we *value* them, or *see* them, as ends. But it appears to be an implication of constructivism that the agent can value only herself as an end.[1] If we can't value others as ends, then we can't treat them as ends. And, since requirements are grounded on our valuings, if we don't value others as ends, we are not required to treat them as ends.

Korsgaard attempts to meet that challenge through her thesis of the publicity of reasons. The thesis of the publicity of reasons is supposed to show that valuing oneself involves valuing others (SN 132).[2] In the previous chapter I elaborated what I think is the correct interpretation of Korsgaard's thesis of the publicity of reasons. I argued that the thesis of the publicity of reasons provides a conceptual framework within which to accommodate the idea of treating others as ends. In the current chapter I explore the way in which that conception of public reasons is supposed to show that valuing oneself involves valuing others.

In order to do that it will help to summarise the particular shape that the moral challenge takes to Korsgaard's own brand of constructivism. We saw in Chapter 3 that according to Korsgaard our practical identities (and the complex relations between them) determine our choice of actions. We choose our actions to preserve our practical identities.[3] Thus put, it seems that whenever our actions involve other agents, those agents will feature in our

[1] As I said in Section 1.1.2. some constructivists don't think that there is one thing which the agent necessarily values. These constructivists don't think that the moral challenge can be met (Street 2008; Lenman 2010; Velleman 2009; Bagnoli 2011), and instead provide a revisionary account of morality (esp. Velleman 2009).

[2] A reminder that by 'valuing' someone I mean valuing as an end (Section 2.3.1. above), and that I am using 'valuing someone' (as an end) interchangeably with 'seeing someone as an end'.

[3] This way of characterising the role of action is developed by Korsgaard under the slogan 'action as self-constitution' (Christine M. Korsgaard 2009).

actions only as means to preserve our practical identities. In order for other agents to feature in our actions as ends, they will have to feature as ends in a practical identity of ours. In that way, in seeking to preserve the practical identity in which others feature as ends, we seek to preserve the others' status as ends for us.

Another basic intuition about morality is that it applies to every agent, and that it does so regardless of their interests. Yet, the particular practical identities are by and large contingent, disposable, and of relative value to the agent. If the status of others as ends were incorporated in one of those practical identities, whether one is subject to morality would depend on whether one has and/or keeps the given practical identity. This would go against the intuition that every agent is subject to morality.

For us to satisfy the intuition that morality applies to everyone, we will have to incorporate the status of others as ends into a practical identity which all agents necessarily have. Only the human identity meets that criterion. This is the practical identity that contains the agent's valuing of herself as an end. If Korsgaard is to accommodate the intuition that morality applies to everyone, she will have to show that the status of others as ends is incorporated not in any practical identity, but in the human identity. And this is what she tries to do. If she succeeds, the human identity will yield the moral identity.

For others to feature in one's human identity in this way – for one's human identity to be one's moral identity – Korsgaard argues, it will have to be the case that valuing oneself 'somehow implies, entails, or involves' valuing others[4] (SN 132). The notion of public reasons was designed forge the connection between valuing oneself and valuing others. Let us then look at the extent to which that connection has been forged, and the extent to which that connection helps establish morality.

4.2.1. Valuing oneself and valuing others

Korsgaard's role for her thesis of the publicity of reasons is explicit enough: through it, she is going to show that valuing oneself involves valuing others (SN 132-136). However, once she has laid out her account of the publicity of reasons, she doesn't explicitly tell us how the

[4] As in the rest of this thesis, and unless otherwise stated, by 'valuing others' I mean valuing others as ends in themselves. I use those expressions interchangeably with 'recognising the normativity of others'.

publicity of reasons is supposed to play that role - how it is supposed to facilitate the required connection between valuing oneself and valuing others. That is, her thesis of the publicity of reasons shows, if successful, how it can be that we see each other as ends. But she doesn't tell us explicitly how our seeing each other as ends follows from seeing ourselves as ends. Nevertheless, I think that we can tease out from her account of the publicity of reasons how that connection is supposed to obtain.

We saw that Korsgaard's thesis of the publicity of reasons accounts for the idea that we see each other as ends by providing a framework in which what makes *my* reasons reasons for me, makes *your* reasons reasons for me. What makes my reasons reasons for me is a given relation I have to myself. There are two aspects of myself involved in this relation: a commitment to doing the right thing (the acting self); and a commitment to finding out what the right thing to do is (the thinking self). I have these commitments because I care about my actions. I care about my actions because it is through my actions that my practical identities are preserved or jeopardised. I care about whether my practical identities are preserved or jeopardised because I value myself as someone who needs reasons to act – as an agent. In other words, I care about whether my practical identities are preserved or jeopardised because I value[5] myself. In short, it is because I value myself that I have an acting self and a thinking self. The two aspects of myself involved in the generation of reasons arise from my valuing myself. This is the way in which what makes reasons for me is a relation I have to myself: a relation of self-value.[6]

This relation of self-value takes a certain shape at the level of the agent's thinking self and acting self as follows. Since the acting self is the manifestation of the agent's commitment to doing the right thing, and the thinking self is the manifestation of the agent's commitment to finding out what the right thing to do is, the pronouncements of the thinking self provide the acting self the opportunity to realise its commitment. That makes the pronouncements of the thinking self normative for the acting self. So, for the acting self to recognise the pronouncements of the thinking self *qua* pronouncements of the thinking self, is for the acting self to recognise the pronouncements of the thinking self as normative. What makes my reasons reasons for me is a relationship between my acting self and my thinking self

[5] As in the previous chapter, unless otherwise stated, by 'valuing' someone, I mean valuing them as ends.

[6] For the sake of concise exposition, I am leaving out issues concerning the standards of correctness of practical identities. These were discussed in detail in Chapter 3, and will recur in different ways in this chapter and especially in Chapter 5.

consisting in my acting self recognising the pronouncements of my thinking self *qua* pronouncements of my thinking self.

If what makes *my* reasons reasons for me is the same as what makes *your* reasons reasons for me, and what makes my reasons reasons for me is the recognition by my acting self of the pronouncements of my thinking self, then what makes your reasons reasons for me must be the recognition by my acting self of the pronouncements of your thinking self. I have access to the pronouncements of your thinking self through your reasons when I recognise them as reasons, i.e. when I recognise them as normative. So in recognising your reasons as reasons, I recognise them as pronouncements of your thinking self. And in recognising the pronouncements of your thinking self, I respond to them just in the same way as in recognising the pronouncements of my thinking self: I take them as the opportunity to realise my commitment to doing the right thing. This makes the pronouncements of your thinking self normative to me.

Just as in the intra-personal case, normativity is generated by the relationship between acting self and thinking self, only according to the thesis of the publicity of reasons, when I share into your reason it is my acting self and your thinking self that are at play. (Notice that the idea that recognising the pronouncements of your thinking self as such, or, equivalently, recognising your reasons as reasons, involves my becoming subject to them, is supported by the constructivist construal of normativity. Since, according to that construal normativity is a force on the will – a distinctive type of force, but a force nevertheless – it couldn't be the case that I recognise the normative standing of anything without experiencing this force, that is, without becoming subject to it. I mentioned this in Section 3.7.1. Notice, too, that as in the intra-personal case, the normative status that the pronouncements of your thinking self have for me comes from the normative status that your thinking self has to me. In other words, it is to your thinking self that I respond to, and only derivatively to its pronouncements.) This is the way in which your reasons become reasons for me; they do so in the same way as what makes my reasons reasons for me – namely, through the recognition by my acting self of a thinking self.

Since, as we have seen, the acting self arises from the agent's self-valuing, there is a way in which my self-valuing leads to your reasons becoming reasons for me. This is the way in which the thesis of the publicity of reasons, if successful, shows that valuing oneself involves valuing others: valuing myself creates my acting self which takes your thinking

self as normative. Once we have that connection between valuing oneself and valuing others, the moral identity follows. Since I can't help but value myself, if valuing myself leads to valuing others, it follows that I cannot help valuing others. If valuing others is contained in my moral identity, it follows that I can't help but have a moral identity. This is the connection between valuing others and valuing oneself – between the human identity and the moral identity – which would establish morality in Korsgaard's account.

In the following sections I explore the extent to which Korsgaard's thesis of the publicity of reasons supports that connection.

4.3.1. From valuing oneself to valuing others: necessity or contingency?

The connection Korsgaard seeks to establish is one where valuing oneself 'implies, entails, or involves' valuing others (SN 132). If, as Korsgaard argues (Section 3.4.1.), agents necessarily value themselves, then it will follow that agents necessarily value others.[7] However, neither of Korsgaard's arguments for the publicity of reasons establishes that connection. The argument from the publicity of language, if successful, shows that I can become subject to your normative status – all the parts and pieces are there; and the argument for the permeability of consciousness shows, if successful, that most of our interactions involve being normative to one another.

But nothing in those arguments suggests that our interactions *must* be like that. Nothing suggests that it cannot be the case that I value myself and not value others. What those arguments would show, if successful, is that that *valuing others* implies, entails, or involves *valuing oneself*. This allows for the possibility that I value myself and yet legitimately not value others. If that is right, it follows that it is not the case that morality applies to all agents. The intuition that morality applies to all agents would not be supported.

Korsgaard seems to acknowledge that her arguments from the publicity of language, and for the permeability of consciousness have not secured the conditional connection which would

[7] Notice that this reasoning applies only if valuing oneself *implies* or *entails* valuing others, but not it valuing oneself *involves* valuing others. For the sake of argument I am focusing on what would be the case if valuing oneself did imply or entail valuing others.

vindicate the intuition that morality applies to everybody, namely, that if one values oneself one values others. Following those arguments, and anticipating a reply similar to mine above, she produces an argument supporting the conclusion that agents necessarily value others. I present this argument below. If her argument were successful, the intuition that morality applies to every agent would be vindicated. I argue, however, that her argument does not support the conclusion that all agents *necessarily* value others, although through a charitable interpretation it can support the conclusion that nearly all agents value others. That is, I argue that although Korsgaard's argument does not give us necessity, it does give us near universality.

Korsgaard's argument for the necessity of valuing others is as follows. She explains that for one to not value others in the way described by her account – for one to not recognise the normative standing of another – one would have to see their reasons as 'mere pressure' (SN 143). That is, if I didn't value you – if I didn't recognise your normative standing – I wouldn't recognise your thinking self. If I didn't recognise your thinking self, I wouldn't recognise your reasons as reasons. If I didn't recognise your reasons as reasons I would see your actions simply as your causal outputs, and I would see *you* entirely in causal terms. But, Korsgaard claims, to see you in that way is impossible (SN 143). If it is impossible for me to see you in solely in a causal way, it follows that it is necessary that I see you as source of normativity for me.[8] We can schematise the argument as follows:

i. In as far as I am aware of you, I must either value you, or see you solely in causal terms.
ii. It is impossible for me to see you solely in causal terms.
iii. Therefore, I must value you (in so far as I am aware of you).

If this argument were sound, then even if Korsgaard doesn't provide an analysis of that necessary connection – even if she doesn't quite explain *why* it should be the case that if one values oneself one must value others – she would nevertheless satisfy the intuition that morality applies to all agents. However, I argue that for all that Korsgaard has said this argument is not sound.

[8] 'It follows' given the wider context of the discussion, of course. The wider context of the discussion supplies the implicit premise that there are two ways in which we might see each other: instrumentally (causally); and normatively (as ends).

I propose that the wider discursive context gives premise i. adequate support. If that is so, the key premise in Korsgaard's argument is ii., namely, that it is impossible for me to see you solely in causal terms. I argue that that premise has not been adequately supported. The only support Korsgaard seems to provide for that premise is the fact that we can't envisage a case where one agent sees another solely in causal terms - a case where an agent hears another's words as mere noise and sees another's reasons as mere pressure. This relies on our presumed inability to imagine an agent who does not recognise the reasons of another as reasons – an agent who sees others and their reasons as nothing more than part of the causal network of the world.

However, since Korsgaard has not provided an analysis of why it should be the case that if one values oneself one *must* value others, we have not been shown that there is something incoherent, or prohibited in some other way, in the idea of an agent who doesn't value others. If the idea of an agent who doesn't value others is not incoherent or otherwise prohibited, we might think that the inability to imagine such an agent simply reveals a conservative imagination. This is not enough to support the premise that it is impossible for such an agent to exist. If that premise is not adequately supported, then Korsgaard cannot give it the pivotal role it has in her argument for the necessity of valuing others.[9] I conclude that Korsgaard is not successful in showing that all agents must value others. Therefore, she is unable to accommodate the intuition that morality applies to all agents.

However, even if our inability to imagine such an agent who sees others solely in causal terms is not sufficient to support the premise that such an agent is impossible, it might be enough to support the premise that such an agent is exceedingly rare. We might suggest that if we can't imagine such an agent it will presumably be partly due to never having encountered such an agent. And if it is the case for most of us that we have never encountered such an agent, it is safe to infer that if there are any such agents, they are rare. If these agents are rare *enough*, if even the most hardened criminals who populate the moral

[9] We could also provide counter-examples to the idea that we must recognise others' reasons as reasons. I see you twitching your foot and I think it is some kind of nervous tick, something short of an action, and thus, something that does not reveal a thinking self. If instead you are twitching your foot as an attempt to draw to my attention that my greatest idol is sitting to my right then I have failed to notice your reason. However, this does not pose such a great threat to morality. I have made a mistake about why you are twitching your foot, but my mistaken explanation about why you were twitching your foot did not eclipse your normative status to me – I continued to attribute reasons to you in regards of at least many of your other actions. What poses the greatest threat to Korsgaard's morality is the possibility of someone failing to recognise the normative status of others altogether.

literature – Hitler, the Mafioso, the psychopath[10] – are not examples of agents who do not recognise others' reasons as reasons, then we might be comfortable enough accepting that it is it exceedingly rare for an agent to not value others, even if it is not impossible.

Replacing the premise that such an agent is impossible with the premise that such an agent is exceedingly rare in Korsgaard's argument above, would yield the conclusion that, with the exception of exceedingly rare cases, all agents value others. We don't get necessity, but we get near universality. Although the intuition that morality applies to all agents is strong, some constructivists have settled for precisely the conclusion that morality might be near universal, but that it is not necessary. If my arguments are correct, then Korsgaard would land in that camp of constructivism.[11]

If we accepted this line of thought then the conclusion which would follow from Korsgaard's attempt to show that valuing oneself involves valuing others would be: in very nearly all agents, the human identity is the moral identity; very nearly all agents have a moral identity; so very nearly all agents are subject to the requirements of morality.

4.4.1. The requirement to value others

The conclusion above states a) that most agents value others; and b) when they do, they are subject to the requirements of morality. However, if we are hoping for a solid foundation for morality, we should hope that our moral account tells us that we *ought* to be moral. That is, we want our moral account to tell us not just that most agents happen to value others, but that those agents *ought* to value others.[12] For Korsgaard to have argued for the requirement *for* morality, she would have to have argued not just that valuing oneself 'implies, entails, or involves' valuing others (SN 132), but that valuing oneself *requires* valuing others. For her to argue that valuing oneself requires valuing others she would have to argue that one is required to recognise others' reasons as reasons.

But Korsgaard has not shown that. Again, neither her argument from the publicity of meaning, nor her argument from the permeability of consciousness, has shown that we are

[10] Bagnoli argues that the Mafioso does not recognise the normative status of others (Bagnoli 2009). I consider Bagnoli's arguments in Chapter 5.

[11] As mentioned in Chapter 1, that camp includes Street, Lenman, Velleman.

[12] A reminder that throughout the text, by 'valuing' I mean valuing as an end, unless otherwise stated.

required to respond to others' reasons. As I said above, all that those arguments show is *how* it happens (argument from the publicity of meaning) and *that* it happens (argument from the permeability of consciousness), but not that it is *required* to happen. So, it looks as though, even if Korsgaard's arguments worked, she can't account for the requirement *for* morality.

However, we might think that Korsgaard's account has the resources to avoid that worry. I explore that possibility in this section. We might think that Korsgaard's work can account for the requirement to value others, and thus to treat them as ends. Our practical identities are normative for us. That means that the values that constitute our practical identities are normative for us. Providing our practical identities are justified, we are required to preserve the valuings that constitute them. Not doing so would be a way of going against their normative status.[13]

For example, if I have a practical identity as a citizen of the UK, I will have values such as voting in the elections, participating in public debates, celebrating achievements by national sports teams or other institutions, following developments in the life of the nation, and so on. As part of my commitment to my practical identity as a UK citizen – that is, as part of having a practical identity as a UK citizen – I am required to maintain those valuings. Failing to do so whilst committed to that practical identity would constitute an affront to the integrity of that practical identity. So, I am required to preserve the valuings constitutive of my practical identity as a UK citizen.

According to Korsgaard, valuing oneself involves valuing others. Valuing oneself is contained in one's human identity. So we might take it that valuing others, when it occurs, is part of one's human identity – it is one of the valuings constitutive of one's human identity. Since one is required to preserve the valuings constitutive of one's practical identities so long as those practical identities are justified, and since one's human identity is necessarily justified, it follows that, *where one already values others*, one is required to value them. So it looks like, despite initial appearances, Korsgaard has the resources to argue for the requirement *for* morality.

[13] And, since the normative status of my practical identities is bestowed upon them by my human identity, to not abide by the obligations they set on me would constitute going against my human identity (Section 3.4.2.). I will say more about this in the next chapter.

However this argument does not work. For this argument to work it would have to be the case that one could not have a human identity (or that one's human identity would be diminished) if one did not value others. But that is not what the argument above shows. All that argument shows is that, *if one values others*, then that valuing of others is constitutive of one's human identity. It doesn't show that one cannot have a human identity without valuing others *if one does not value others*. So, if you value others as part of your human identity, that valuing is constitutive of *your* human identity (and as such you will be required to value them); and if you don't value others, then valuing others is not part of *your* human identity (and so you are not required to value others). In other words, valuing others is not necessarily constitutive of the human identity as a type; but where it is part of a particular human identity, then it is necessary for that token human identity.

Let me illustrate this with an example. Take the practical identities of being a vegetarian and of being a UK citizen. Let's assume that it is necessarily the case that if you are a vegetarian you don't eat animals, and that if you are a UK citizen you are committed to abiding by the constitution (or what goes as a constitution on these shores). These are valuings that you have to have in order to count as having those practical identities. Since I have established that one is required to endorse the valuings constitutive of one's practical identities in so far as one holds them as practical identities (and those practical identities are justified), then *everyone* who has the practical identity of being a vegetarian or of being a UK citizen is required to endorse not eating animals or abiding by the UK constitution.

But suppose that as part of your practical identity of being a vegetarian you elect to donate money to animal welfare charities; and that as part of your practical identity as a UK citizen you elect to take a day off work every year to watch the opening of Parliament. These valuings are necessary for *your* practical identities as a vegetarian and as a UK citizen. If you acted against them (whilst you have them) you would be jeopardising their integrity. But those valuings are not necessary for anyone else to have those practical identities (unless they do have those valuings). Someone can have a practical identity as a vegetarian or as a UK citizen without donating money to animal charities, or watching the opening of Parliament every year. And for such a person not giving money to animal charities or not watching the opening of Parliament would not constitute an affront to her practical identities. As such, that person would not be required to give money to charities or to watch the opening of Parliament every year.

For the conditional *if valuing others is constitutive of the human identity, then everyone who has a human identity must value others* to hold, it would have to be the case that valuing others is constitutive of the human identity in the same way that not eating animals is constitutive of the practical identity of being a vegetarian, or abiding by the UK constitution is constitutive of the practical identity of being a UK citizen. But the argument above does not show that. All that the argument above shows is that valuing others is part of one's human identity in the same way as donating money to animal charities or watching the opening of Parliament are constitutive of the practical identities of being a vegetarian or of being a UK citizen. So, just as one is required to give money to animal charities or to watch the opening of Parliament only if one values those things as part of one's practical identities, one is required to value others only if one already values others as part of one's human identity. This then is why all the argument above would show is that morality is required only for those who happen to value others – for those who are already part of morality.[14]

[14] It might be argued that there is a more important implication to be drawn from the conclusion that not everyone is required to value others as ends, and so that not everyone is required to reflect the value of others in their actions. That is what that conclusion says about the nature of practical rationality. Part of the aim of the account of public reasons was to vindicate Kant's argument for his Formula of Humanity. Kant's argument was that, since one necessarily represents oneself as an end in itself, and so do all other persons, the law that follows from that is that one is to act in such a way that one treats oneself and others as ends in themselves. As noted in Chapter 2, this argument is widely thought to commit a non-sequitur. Whilst we might grant that if one values oneself one is to reflect that in one's actions, it doesn't follow from other's valuing themselves that one must also treat them in a way that reflects their value. Yet that is what Kant's argument purportedly establishes. Korsgaard holds that if it is the case that valuing oneself leads to valuing others, then Kant's argument follows a clear logic. She sets out to show that it is the case that valuing oneself leads to valuing others through her account of the publicity of reasons. However, since Korsgaard has not established that valuing oneself always, or necessarily, involves or requires valuing others, but only that in most cases it does, we might think that she has fallen short of lending Kant's argument the support it needed. This is because whilst Kant's intended conclusion is categorical and universal, Korsgaard's is hypothetical and not-universal. As such, it does not count as the a priori deduction of the moral law which Kant sought.

However, there is a plausible way of interpreting Kant's argument in light of which Korsgaard has supplied the support needed. This way of interpreting Kant's argument sees it as claiming that, since one must think of oneself as an end in oneself, and - and this is where the distinctiveness of this interpretation resides - one *recognises* that others must also see themselves in this way, then one must treat oneself and others as ends in themselves. If recognising that another must treat herself as an end in itself is tantamount to recognising that she is a person, the Formula of Humanity does not say to treat oneself and every other person – regardless of whether one sees them as persons – as ends in themselves. Rather it says to treat oneself and those others *who one recognises as persons* as ends in themselves. If we can infer from Korsgaard's discussion that to recognise someone as a person is to respond to their reasons, then it will be the case that valuing oneself as an end and recognising those others who also value themselves as ends will *lead* to one treating oneself and others as ends in themselves. This, as it stands, does not fully support Kant's argument, for it only says what happens, whilst Kant's argument yields an imperative: one is to act in such a way that oneself and others are treated as ends. However, since we have also seen that the requirement to treat others as ends applies to those who already treat others as ends, then we are not saying simply what happens, but what ought to happen. Kant's argument for his Formula of Humanity would then be fully supported, as far as our discussion of Korsgaard goes thus far. I will argue later that Korsgaard's account cannot support the Formula of Humanity.

Korsgaard's attempts to establish morality by establishing a link between valuing oneself and valuing others don't show that that link is either causally necessary, or normatively necessary. That is, her arguments do not show either that, necessarily, if one values oneself one values others; nor that, necessarily, if one values oneself one *ought* to value others. However, it looks like Korsgaard's account has the resources to establish that in very nearly all cases, if one values oneself one values others and that when that is the case, one is required to value others. In short, it looks like Korsgaard has established that in very nearly all cases, one is required to value others. [15]

In the next chapter the question of whether agents are required to value others is put under pressure from another angle. I examine the relation between Korsgaard's account of the publicity of reasons and the Kantian background within which she works. I will argue that the thesis of the publicity of reasons is in conflict with the requirements of autonomy. Since the notion of autonomy is implicit in Korsgaard's notion of the human identity, it follows that the publicity of reasons is in conflict with the human identity. This is but a guise of the original moral challenge to constructivism as set out in Chapter 2: whether it can accommodate morality given its account of normativity. I will argue that it can't. If the publicity of reasons is in conflict with the human identity, and our moral account is based on the publicity of reasons, it follows that one is required to *not* be moral.

[15] It might appear that we can apply the same considerations to the moral identity as in the previous footnote I suggest we might apply to Kant's argument for his Formula of Humanity, and thus end up with a moral identity even in those who do not recognise the value of others. However, that strategy would be less applicable here, for Korsgaard's account is more substantive than Kant's. She sets out to show that we do value others.

- 5 -

Public reasons and autonomy

5.1.1 Korsgaard and Kant[1]

Korsgaard's attempt to establish a constructivist morality is also supposed to be a way of improving Kant's account of morality. I mentioned in Section 2.4.1., that if Korsgaard succeeds in showing that valuing oneself involves valuing others, she will also vindicate Kant's argument for the Formula of Humanity. As we also saw in Section 3.4.3., one of the aspects of Kant's theory which Korsgaard thinks needs amending is the relation between the categorical imperative and the moral law. Korsgaard's approach to Kant's argument for the Formula of Humanity, and to the relation between the categorical imperative and the moral law, are closely related. So much so that if she succeeds or fails to account for one of them, she correspondingly succeeds or fails to account for the other. We shall see this later. Now, I want to spell out in a bit more detail Korsgaard's take on the categorical imperative in contrast with Kant. I will then look at the implications of Korsgaard's approach to other aspects of Kant's moral theory, specifically, his notion of autonomy.

Whilst Kant thought that the categorical imperative and the moral law are the same thing, Korsgaard thinks that they are not. Kant thought that the categorical imperative and the moral law are the same thing because he thought that the scope of application of the categorical imperative includes all agents. On that view my application of the categorical imperative when trying to decide whether to ϕ, tests whether I can will that the maxim of that action be a principle to *all* agents. To base my decision on whether I can will that my maxim be a principle to all agents is to treat all agents as ends in themselves. To treat all agents as ends in themselves is what the moral law requires. So the requirement of the moral law is the requirement of the categorical imperative.

[1] In this section I produce an outline summary of Korsgaard's position, bringing to the fore its connection with Kant. For a fuller version of her position see Sections 2.4.1. and 3.4.3.

Korsgaard's account of practical reasoning yields a different conclusion about the scope of the categorical imperative. For Korsgaard the normativity of practical reasoning comes from oneself. We are essentially and primarily self-valuing creatures, and as such we seek for our self-valuings to be preserved through action. Those self-valuings take the shape of practical identities, and they are ultimate grounded on the human identity. The human identity, then, is the only fundamental and necessary practical identity. Practical reasoning is the way in which we ensure that our practical identities are preserved through action. So the normativity of practical reason is grounded in the agent's self-valuing - ultimately in the agent's self-valuing as an agent, in the human identity.

Since the categorical imperative is the principle of reason, the normativity of the categorical imperative is grounded in the agent's self-valuing. According to Korsgaard's account, for the agent to apply the categorical imperative is to seek to determine whether she can will that her maxim be a principle for *herself*. The scope of application of the categorical imperative is the agent alone because the categorical imperative is the way in which the agent ensures that her choice of action preserves herself under a certain conception - ultimately as an end for herself. This, as we saw, is the way in which Korsgaard crafts her version of the main tenet of constructivism, namely, the idea that normativity originates in oneself.

As we have also seen, Korsgaard does think that the categorical imperative can yield the moral law. The categorical imperative yields the moral law when it applies to all agents, that is, when in the process of reasoning the agent applies the categorical imperative to all agents. If the agent applies the categorical imperative to herself because she values herself, then for the agent to apply the categorical imperative to all agents she will have to value all agents. So the categorical imperative yields the moral law when the agent values all agents as ends in themselves.

Korsgaard captures the idea that the agent sees all agents as ends in themselves through the notion of the moral identity. The moral identity is the practical identity in which the agent sees herself as an end, but she also sees the other agents as ends. In other words, it is the practical identity in which the agent sees herself as one amongst other ends (SN 132). For it to be the case that actual agents apply the categorical imperative to all agents – that is, for it to be the case that the categorical imperative takes the shape of the moral law – it will have to be the case that the agent does see all agents as ends in themselves, that she does have a

moral identity. Korsgaard tries to show that the agent does have a moral identity through her account of the publicity of reasons.

The categorical imperative is connected to several other critical aspects of Kant's moral philosophy. As such, the adjustments which Korsgaard makes to the notion of the categorical imperative might have repercussions for the broader Kantian moral picture. One such connection is with autonomy. In the next section I argue that this connection has been disrupted by Korsgaard's adjustments to the notion of the categorical imperative. To show how, I begin by presenting Kant's account of autonomy, and its connection to the categorical imperative.

5.2.1 Kant and autonomy[2]

According to Kant, autonomy is the property of the will by which the will determines itself.[3] Kant develops that thought to account for the idea that agents are the authors of their actions, as opposed to actions being something that happens to agents (as Hume's account of action *prima facie* suggests). Actions, according to Kant, are not *solely* events determined by causal relations. There is a way in which they are. If a scientist were to look at your brain when you make decisions, she might be able to see the sequence of events in your brain that culminate in your action. She might even be able to predict what this sequence of events will be, and consequently what your action will be. So, from this point of view, from a third personal point of view, your actions are part of the causal fabric of the world.[4] But there is another point of view from which things are not that way. That is the practical point of view.

The practical point of view is the point of view occupied by the agent when she deliberates. For the deliberating agent things are very different than they are for the scientist monitoring her brain. One way in which deliberation is different for the agent than for the

[2] What follows is a sketch of Kant's characterisation of autonomy in GW 4:440-4:463.

[3] We have already seen various elements of Korsgaard's work based on Kant's (her use of the categorical imperative; her attempt to vindicate the Formula of Humanity). We are about to see that Kant's account of autonomy is the inspiration of Korsgaard's account of the normative mind, as I presented it in Section 3.3.1. This will be of relevance for my argument in Section 5.5.1.

[4] Also, and less fancifully, most of the things you do involve making changes to your body in order to make changes in the world. The causal aspect of your actions is evident there too.

scientist is that, whilst for the scientist the agent's actions are the result of a series of causal events, for the agent they are the result of her practical deliberation alone. That is, in setting out to decide whether to ϕ, the agent presupposes that her decision will be determined by herself through a process of deliberation. The presupposition that one will reach one's decision for action oneself through one's deliberation is not a belief which plays the role of a premise in one's deliberation. Rather, it is a belief to which one is implicitly committed to in deliberating.

To say that your conclusion won't be determined by anything other than your own deliberation implies that psychological states – states such as desires, attractions, aversions, and the like – won't determine your conclusion. Deliberation, on this view, is not a matter of, say, your desires battling one another until the victorious one leads you to action. Desires may be present in deliberation, to be sure, but not as determinants of the agent's conclusion. Rather, desires are present in deliberation as objects of the agent's reflection.

For example, if you are wondering whether to watch the opening of Parliament or not, your desires will play a role – whether and how much you want to watch it, or to do something else instead will be objects of your deliberation. But you don't assume that those desires will determine what you decide to do. If you assumed that your desires will determine what you decide to you, there would be no point in your deliberating about what to do – your decision won't be delivered by your reasoning. Even if you decide not to watch the opening of Parliament because you would much rather go to the theatre to see a performance which takes place at the same time, that decision won't be the result of your desire, but of your deliberation, even if your deliberation included considering your desire to go to the theatre. Equally, it wouldn't make sense for you to deliberate about whether to watch the opening of parliament if you believed that the decision would be planted in your mind by a hypnotist, rather than formed by your own deliberation. In as much as one deliberates, one is committed to the idea that one will reach one's decision through one's reasoning alone.

This is a description of what Kant calls the negative aspect of autonomy: the way in which when deliberating we cannot but do it under the idea of freedom – freedom from determinants other than ourselves.

This negative aspect of autonomy raises a question that demands a positive account. If deliberation is done through principles, as Kant thinks it is, and in deliberating the agent is committed to reaching her conclusion through her deliberation alone, then the content of

those principles cannot refer to anything outside her reason or will.[5] If they did, those principles would be causal. If the principles of her deliberation were causal, it would not be the case that the agent determines her conclusions through reason alone. If the content of the principle of reason cannot refer to anything outside of reason, then it must refer to itself. That is, it must say to just be a principle – a law. Laws, or principles, are universal. So, if the principle of reason says to be law, it says to be universal.

When applied to specific instances of reasoning, that law says that my maxims must be fit as a principle or law. In other words, the principle of reason requires that one's maxims be universalisable. But that is just the application of the categorical imperative: to test whether one's maxim is fit as a principle – i.e. whether I can will its universalisation.

That, then, is the connection between autonomy and the categorical imperative in Kant's work: autonomy is the way in which the agent is normative to herself, the way in which she determines the conclusions of her deliberation; and the agent determines the conclusions of her deliberations through the application of the categorical imperative. In other words, the agent expresses her autonomy though the application of the categorical imperative.

Now, although all agents are autonomous – all agents, in as much as they deliberate at all must assume their own autonomy – it doesn't follow that all agents *express* their autonomy all the time. That is, it doesn't follow that all agents always apply the categorical imperative fully in their deliberation. When they don't, they instantiate heteronomy. This is how Kant puts it:

> … if the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law … heteronomy always results. The will in that case does not give itself the law; instead the object, by means of its relation to the will, gives the law to it. This relation, whether it rests upon inclination or upon representation of reason, lets only hypothetical imperatives become possible: I ought to do something because I will something else (Kant 1998, p.4:441).

That is, if when looking to decide whether to φ my deliberation is motivated not by wanting to do the right thing, but by something else – e.g. by wanting to satisfy my desire, or gain acceptance, or to avoid confrontation or hardship – then I am not allowing my will to impose its own law. Instead I am allowing myself to be led by something extraneous to it. As such, my action or decision is heteronomous.

[5] Kant equates practical reason with the will (Hill 1989).

Examples of heteronomous decisions are deferring to 'the dogmas of the Church, to the edicts of rulers, to immediate inclination, or to the will of the majority' (O'Neill 2003, p.9). Korsgaard herself uses the character Harriet, from Jane Austen's *Emma*, to illustrate heteronomous actions[6] (Korsgaard 2009, p.162). Harriet, who feels inadequate in many respects, governs her decisions by what Emma thinks she should do, or by what she thinks Emma thinks she should do.

Harriet acts heteronomously in those cases in that she allows herself to be determined by someone else's will: Emma's. Harriet's actions are heteronomous even if Emma's decisions are better than Harriet's, and even though in letting herself be led by Emma's decisions, Harriet is letting herself be led by Emma's *reason*. What matters is that it is not her own reason that she follows. Since it is not her own reason that she follows, she violates the negative aspect of autonomy. Furthermore, in allowing her choices to be decided by someone else, Harriet is not guiding herself by the universalisation requirement; she is not seeking to make the right decision, instead she is perhaps seeking to avoid the possibility of making a mistake, or of contradicting Emma, or she might simply see Emma as having greater authority over herself than herself.[7] In any of those cases, she is failing to reach her conclusions through the categorical imperative, and thus she is violating the positive aspect of autonomy.

If the agent fails to express her own autonomy then she positions herself against her own normative status to herself. That is because in the very act of engaging in deliberation the agent is committed to her own autonomy. In being committed to her own autonomy, the agent is committed to expressing her own autonomy. And in being committed to expressing her own autonomy, the agent is committed to reaching her own conclusion by ensuring that her maxims are fit as principles. That is, in being committed to expressing her own autonomy the agent is committed to reaching her conclusion through the application of the categorical imperative. But if *whilst she remains thus committed* she allows the conclusion of her deliberation to be settled by something other than the rightness of her principles of choice, then she is raising herself against her own autonomy. In other words, she is going

[6] Strictly speaking, as O'Neill makes clear, the expression of autonomy, or heteronomy, pertain to choosings, rather than actions (O'Neill 2003). However, when I talk about heteronomous actions, or actions which express one's autonomy, I assume that actions have those properties in virtue of the choosings from which they issued. That is, an action expresses one's autonomy or is heteronomous depending on whether the principle on the basis of which that action was chosen was expressive of the agent's autonomy or was heteronomous.

[7] Hill (Hill 1973; Hill 1983) explores the relation between autonomy and self-respect commonly understood.

against the very process that she has initiated and remains committed to: practical reasoning. So, heteronomy is *wrong* by the agent's own lights.

A note about how I shall be using the notion of heteronomy in the following chapters. Descriptively speaking, autonomy contrasts with heteronomy. Something is autonomous if it rules itself; something is heteronomous if it is ruled by something other than itself. But, normatively speaking, autonomy contrasts with lack of autonomy – *not* with heteronomy; and heteronomy contrasts with the expression of autonomy. Since autonomy is the property of will by which it governs itself, and it is in the nature of the will to govern itself, it follows that everything with a will is autonomous. That means that all persons are autonomous. Since only autonomous beings can express their autonomy, or fail to express it, only autonomous beings can be heteronomous. Only autonomous beings can *choose* to act on a principle which is not universalisable, and thus act heteronomously, or on a principle which is universalisable, and thus express their autonomy. I will use the notions of autonomy and heteronomy in their normative senses. In this I follow O'Neill (O'Neill 2003, p.9).

We have seen, then, that according to Kant, the agent's autonomy is expressed by her will, or reason, governing itself. The will governs itself through the application of the categorical imperative. Failing to make one's decision through the application of the categorical imperative – viz. deciding what to do on the basis of something other than the suitability of the agent's maxim as a principle of choice – is an instance of heteronomy, of being governed by a law which is not one's own. Since Kant thought that the categorical imperative is both the principle of reason and the moral law, he thought that the expression of autonomy made the resultant action both the agent's own, and morally worthy. This is how we obtain one of the most inspiring ideas in Kant's moral philosophy: that to act morally is to act freely, or autonomously.

If Korsgaard is right that reason's own principle is not the moral law – if her distinction between the categorical imperative as the principle of reason and the categorical imperative as the moral law is correct – then we have to reconsider the relation between autonomy and the moral law, and thus between actions expressive of one's autonomy and morally worthy actions. Since Korsgaard agrees with Kant that the categorical imperative is the principle of reason, then the connection between autonomy and the categorical imperative as a

principle of reason will remain. So, if Korsgaard is right, it will remain the case that to act autonomously is to act freely.

But, since according to Korsgaard, the principle of reason is the moral law only when it issues from the moral identity, the agent's autonomy will be expressed through the application of the moral law only when the agent deliberates from her moral identity. In other words, according to Korsgaard, to act autonomously is to act morally only if one's action has been chosen from one's moral identity. As we have seen, Korsgaard argues that the moral identity is formed through the publicity of reasons: I recognise your normative status by responding to it, and that involves my conceiving of you as having a normative standing equal to my own. However, as I am about to argue, Korsgaard's account of the publicity of reasons, if successful, rules out the expression of the agent's autonomy, and, in doing so, it severs Kant's connection between autonomy and morality.

5.3.1. Public reasons and the expression of autonomy

I begin this section by highlighting the similarity between the publicity of reasons and heteronomy. For this we need look no further than Korsgaard's own characterisation of the way in which reasons are public – the way in which we are normative to one another. The language she uses here is evocative of heteronomy. The fact that we need reasons to act, and that we can easily intrude into each other's consciousness, she says, both enables and *forces* the extension of reasons across persons.[8] Given the permeability of consciousness, just by talking to you, I can *obligate* you, I can *force* you to think (SN 138, 139, 140). This way of describing the phenomenon of the publicity of reasons is not to be construed as saying that one is forced to act through brute force. If that were the case, the result would not be an *action*, properly speaking, as actions require that they have been *chosen* by the agent. Rather, when your reasons are reasons to me, they exercise a compelling grip on me. But I can choose whether to give in to that grip or not (I'll discuss this further below). Giving in to the grip which your reasons have upon me very much seems like allowing myself to be governed by your will. But this is just what Harriet does, and what it is for actions to be heteronomous.

[8] '[…] what both enables us and forces us to share our reasons is … our *social nature*' (SN 135). Emphasis original.

These considerations are intended to draw attention to the similarities between the descriptions of heteronomy and of the publicity of reasons. They are, thus, superficial similarities. However, I will now argue that the similarity between heteronomy and the publicity of reasons does not end at the superficial level. I argue that acting on another's reasons in the way advanced by the thesis of the publicity of reasons, is an instance of heteronomy. As such, the publicity of reasons and the expression of the agent's autonomy preclude each other.

To properly examine the idea that the publicity of reasons precludes the expression of the agent's autonomy it will be helpful to lay out clearly the main concepts in play. Some of these concepts have already been used throughout this thesis, and their meaning there is no different from their meaning elsewhere. Nevertheless, in the following discussion the differences and relations between them are of particular importance. The main concepts are: seeing someone as an end; treating someone as an end; and becoming subject to another's normativity. I spell them out in turn.

- *Seeing someone as an end.* I explained in Section 2.2.1. that to *see* someone as an end is to see them as setting constraints on what you may or may not do to them regardless of how it impacts on your ends (or on someone else's ends - I'll take this as given in what follows). It is to see them as having a normative standing for you. We have seen that, for the constructivist, normativity is experienced as a claim on the will (Section 1.1.2.). So, for someone to have a normative standing for you, is for that person to make a normative claim on your will.[9] To *see* someone as an end, then, is to be aware that they make a claim on your will. And for someone to make a claim on your will is for you to be subject to their normativity.

- *Treating someone as an end.* To treat someone as an end is to govern your actions by the normative claims they make on you, *because they make those claims*, regardless of how that impacts on your own ends. That is, if I make a claim on your will for you to φ, you will treat me as an end if you φ because that is what my claim in your will says to do. You will not treat me as an end if you don't φ, or if you φ but not because that is what my claim on your will says to do.

[9] In saying that A *makes* a claim on one's will I don't mean that A makes that claim intentionally. I explained in Section 2.5.3., that the claim which another makes on one's will does not have to be intentional.

If you are to do what my claims on your will say to do *because those claims say so*, you will have to be aware of those claims. Since being aware of my claims on your will is to see me as an end, for you to *treat* me as an end, you have to *see* me as an end. But it doesn't follow that if you *see* me as an end, you will *treat* me as an end. If you are my friend presumably I see you as an end. But if I inform your stalker of your whereabouts in exchange for some cinema tickets I am treating you as an end.

- *Becoming subject to another's normativity*. I said above that to see someone as an end is to be subject to their normativity. By '*becoming* subject to another's normativity' I employ the ordinary sense of 'becoming', to mean going from not being subject to another's normativity to being subject to it. Or, in other words, from not seeing another as an end, to seeing them as an end.

- Finally, there are several different expressions which I shall use to mean the same thing. They are, and their connection is, as follows.

  According to the thesis of the publicity of reasons, to see another's reasons as reasons is to see the other as a reason-giver, as a source of normativity. Since, as I have said above, to recognise the other's normative status is to be subject to it, to see another's reasons as reasons is to be subject to them. Since Korsgaard's constructivist construal of normativity has it as a kind of claim on one's will, to be subject to another's reasons is for those reasons to make a claim on one's will. So, the expressions 'being subject to another's normativity', 'being subject to another's reasons', 'valuing another as an end', all mean seeing another as an end.

  The relation between those different expressions translate exactly when we are talking about *treating* someone as an end, or about becoming subject to another's normativity. That is, 'treating another as a source of normativity', 'treating another's reasons as reasons' and so on, mean treating another as an end. And 'becoming subject to another's normativity', 'becoming subject to another's reasons' and so on mean coming to see another as an end.

With all that terminology in mind, in what follows I argue that *becoming* subject to another's reasons is not something the agent decides, and so it is not a vehicle through which the agent could express her autonomy or fail to do so. This is consonant with the

critique of Korsgaard's arguments for the publicity of reasons developed above. However, the agent does decide whether to *treat* another's reasons as reasons. Deciding whether to treat another's reasons as reasons is then something through which the agent expresses her autonomy, or fails to do so. I argue that the agent cannot autonomously decide to treat another's reasons as reasons. I argue that deciding to treat another's reasons as reasons can only be an instance of heteronomy.

5.3.2. Becoming subject to another's reasons

As we saw in sections 3 and 4 above, Korsgaard's arguments for the publicity of reasons show only how it is that the publicity of reasons *can* obtain amongst creatures like us, and that it *does* obtains amongst us. Korsgaard appeals to Wittgenstein's argument for the publicity of meaning to illustrate her account of how the publicity of reasons can arise between us. She then applies the idea of the publicity of reasons to analyses of our interactions to show that it *does* obtain amongst us. It was one of the observations in section 4.3.1. above that Korsgaard does not argue that one is *required* to become subject to another's reasons. (We then saw that we might find a way of arguing that one is required to *remain* subject to another's reason. That conclusion will be challenged later in this chapter, but in any case we can ignore it at present, as our concern here is with *becoming* subject to another's reasons.)

Korsgaard does not argue that we are required to become subject to others' reasons, because she doesn't believe we are. When we talk about a *requirement* to become subject to others' reasons, we mean a rational requirement. Korsgaard believes that reason cannot produce such a requirement. That she believes that is clear from her analysis of the objection commonly raised to Kant's argument for the Formula of Humanity, and it follows from her account of the way in which reasons are public.

The argument for the Formula of Humanity can be distilled to this: Since one must value oneself as an end in oneself; and others must equally value themselves as ends in themselves; then one is to always treat others and oneself as ends in themselves. The conclusion of this argument has two parts: simultaneously one is to treat oneself as an end,

and to treat others as ends. But critics argue that only the first part of this conclusion is adequately supported.[10]

As I explained in Section 2.3.1., it doesn't seem to follow from my recognising that others must see themselves as ends in themselves, that I must treat them as ends in themselves. Appeals to consistency yield at most the requirement that I recognise that others are to treat themselves as ends. Appeals to self-interest defeat the object of morality. Korsgaard agrees with this critique. Specifically, she agrees that there isn't a *reason* to come to value others as ends (SN 134), and thus to treat others as ends.

What she disagrees with is the supposition that the link between valuing oneself and valuing others consists of a reason. That is, she thinks that it is an error to look to reasoning as a way to find the connection between valuing oneself and valuing others. To think of reasons in that way – to think that *my* reasons can lead me to value you as an end – is to think that reasons are private (SN 133). But private reasons cannot lead one to value another as an end, just as the criticism of the argument for the Formula of Humanity outlined above and developed more fully in Section 2.3.1. shows.

Korsgaard's answer to this problem is precisely to argue that reasons are public in the sense that their normativity 'spreads' across agents. We are normative to each other just because we are permeable to each other's reasons - to each other's normative status. We are permeable to each other's normative status in virtue of the relational nature of normativity. And normativity is relational in virtue of the reflective nature of consciousness. The argument from the publicity of language highlights the relational structure of normativity, and the argument for the privacy of consciousness illustrates the idea that our consciousness is permeated by each other's normative status.

So, Korsgaard's solution to the problem of bridging the gap between valuing oneself and valuing others is to argue that there is no gap to bridge. This is the way in which, according to Korsgaard, 'Human beings are social animals in a deep way … the space in which … reasons exist … is a space that we occupy together' (SN 145); '[o]ur social nature is deep in the sense that it is in the nature of our reasons that they are public and shareable' (SN 136); and '… what both enables us and forces us to share our reasons is, in a deep sense, our *social nature*' (SN 135).

[10] As I noted earlier (Chapter 2) some commentators think that the argument does not support even the first part of the conclusion, e.g. (S. L. Darwall 2006; S. L. Darwall 2009).

Before moving on, I should note two things about Korsgaard's position. One is that her account remains rationalist in as much as it is the nature of *reasons* that enables their sharing. But it is not rationalistic, as it rejects the idea that reason*ing* can lead one to value another as an end. The other thing to note concerns the criticism to the argument for the Formula of Humanity rehearsed above and more fully in Section 2.3.1. That criticism is used to show that one is not *required* to come to value others - to *become* subject to another's normativity - but they leave open the possibility that it is *permissible* that one does so. However, Korsgaard's assessment of the conception of reasons underlying those criticisms implies that to apply the concepts of rational requirement and rational permissibility to the idea of one coming to value others is to commit a category mistake. Reasoning can't lead you to value another as an end.

I conclude, then, that it follows both from Korsgaard's motivation for developing her account of the publicity of reasons, and from her account of the publicity of reasons itself, that becoming subject to another's reasons is not something that the agent decides. Rather, it is something that is part and parcel of our social and rational nature. If becoming subject to your reasons is not something I decide, then it is not something through which I can express my autonomy, or fail to do so.

5.3.3. Treating others as ends

So I become subject to your reasons not by deciding to do so, but through the permeability of my consciousness - I simply recognise the claim that you make on my will as a claim. But, once that claim is on my will I have to decide whether to endorse its status as a claim from you, or to resist it. That is, I have to decide whether to do what your claim on my will says to do *because your claim says to do it*, or not. To endorse the status of your claim is to treat you as an end; to resist it is to not treat you as an end.[11] I argue that my decision to treat you as an end can only be heteronomous.

[11] There is a third logical possibility: that I treat you as a non-end. However treating you as a non-end is a way of not treating you as an end. And for my purposes the dichotomy [treating you as an end v ¬(treating you as an end)] is sufficient.

As I have mentioned repeatedly, constructivists look at normativity as a claim on the agent's will.[12] Desires and pro-attitudes in general are claims on one's will too. However, normativity is supposed to be distinct from others claims on one's will due to its origin. Normativity originates in the agent herself and is directed to the agent herself. That is, the difference between the claim on my will which my *decision* to φ makes, and the claim on my will which my *desire* to φ makes is that in the case of my decision I have put that claim on my will myself – I have expressed my autonomy – whilst in the case of my desire, it has appeared independently of my volition.

But if the normative claim which your reasons place on my will is not a decision of mine – if it appears independently of my volition – your reasons are on a par with my desires. They both make a claim on my will independently of my deliberation, they both are pressing me to take some specific actions, and in both cases I have to decide whether to do as they say or not. The question I now ask is, can I autonomously decide to do as your reasons say to do?

My answer is that I can, just in the same way as I can autonomously decide to do as my desires say to do. And just as in the case of desires, I can also decide to do what your reasons say to do heteronomously. However, I argue that for my decision to do as your reasons say to do to be expressive of my autonomy, I have to not *treat* you as an end. Conversely, if my decision to do as your reasons say to do involves treating you as an end, my decision will be heteronomous.

To explain that, I shall first present the symmetry between on the one hand autonomously/heteronomously deciding to do as my desires say to do, and on the other hand autonomously/heteronomously deciding to do as your reasons say to do. With this symmetry clearly in mind I shall be able to show why an autonomous decision to do as your reasons say to do is incompatible with treating you as an end.

Let's look at desires first. Suppose I become aware of a desire to eat a piece of chocolate. I can decide whether to eat the chocolate in one of two ways. I might decide to eat the chocolate just on the basis that I desire to do so, or I can decide to eat it because I decide that it is the right thing to do. For me to decide to eat the chocolate just on the basis that I desire to do so, is for me to have reached my decision on the principle that I do whatever I

[12] I explained that particularly in Section 1.1.2.

desire. But since my desires are claims on my will that have not been issued by my volition, they are not expressions of my will. If my desires are not expressive of my will, the principle that I do whatever I desire allows my decisions to be determined by something outside my will. As such that principle would contradict my own normative status to myself and thus could not pass the categorical imperative test. Therefore, my decision to eat the chocolate just because I desire so would be heteronomous.

By contrast, for me to decide to eat the chocolate because it is the right thing to do, is to decide to eat it because it passes the categorical imperative test. Perhaps I have been working hard and think that to treat myself to the chocolate would do me good. I judge that a principle containing these considerations is universalisable, and on that basis, I decide to eat the chocolate. In this case the fact that I have a desire to eat the chocolate is considered in deliberation, but it doesn't determine my decision. My decision is determined by my judgement that eating the chocolate given the circumstances would be the right thing to do. In this case my decision to eat the chocolate would express my autonomy.

So my decision to do as my desire says to do can take one of two routes: I can allow my desire to determine my decision, and thus be heteronomous; or I can ensure that doing what its content describes (eating the chocolate) is the right thing to do, and thus express my autonomy.

The same two routes are available for my decision to do what your reasons say to do. Suppose you ask me not to park my car outside your door. Your request registers with me as a claim on my will – a claim to do as your reason says. This is analogous to having a desire for chocolate: in both cases I am aware of a claim on my will which has not been placed there by myself, i.e. which is not a decision I have made. As with my desire, I have to decide whether to do as that claim on my will says to do. In this case I have to decide whether to do as your reasons say to do. As in the case of desires, I might decide to do what your reasons say to do because they are your reasons; or I might decide to do what your reasons say to do because I conclude that doing what they say to do – in this case not parking my car outside your door – is the right thing to do.

For me to decide to not park my car outside your door *just because that is what your reasons say to do*, is for me to have reached my decision on the principle that I do whatever your reasons say to do (whatever your reasons say to do *to me*, of course). But since your reasons, when they are reasons for me, are claims on my will which have not been issued by

my volition, they are not expressions of my will. If your reasons, when they are reasons for me, are not expressions of my will, the principle that I do whatever your reasons say to do to me allows my decisions to be determined by something outside my will. As such that principle would contradict my own normative status to myself, as it wouldn't pass the categorical imperative test. Therefore, my decision to not park my car outside your house just because you ask me to, would be heteronomous.

Instead of deciding not to park my car outside your door just because you have asked me to, I could decide to do it because I conclude that not parking my car outside your door is the right thing to do. For example, I might reason that abiding by your request will be conducive to establishing cordial relations with you, which is something that will make my life easier; or I might not want to upset you, as that upsets me; or I might simply conclude that there is a better parking spot available.

It is plausible that each of those considerations could be captured in a principle which would pass the categorical imperative test, and for the sake of argument I propose that we accept that they can. In this case, the fact that you have asked me not to park my car outside your door – that is, the fact that you placed a certain claim on my will – has been considered in my deliberation, but has not determined the outcome of my deliberation. The outcome of my deliberation has been determined by my judging that not parking outside your door would be the right thing to do, given any of the considerations outlined above. In this case, my decision to not park outside your door will be expressive of my autonomy.

As in the case of desires, I can allow my decision to be determined by a claim on my will, namely, your request, and be heteronomous; or I can make my decision on the basis that it is the right thing to do, and thus express my autonomy. However, I argue that to treat you as an end is to act heteronomously, and, conversely, to act autonomously requires that I don't treat you as an end.

According to the thesis of the publicity of reasons, I am supposed to treat your reasons – your authority – as an end for me. In this way, once I recognise your normative status for me – once your reasons are reasons for me – I am to do as you say because you are normative for me (there is an important condition attached to this which I will consider shortly). In other words, I am to do what you say just because you say it. But as I have shown above, that is an instance of heteronomy. It is an instance of heteronomy because your normativity for me is a claim on my will which has not issued from my will. To decide

to do what you say just because you say to do it is to decide to do what an extraneous claim on my will says to do just because it is a claim on my will, and despite the fact that it is a claim extraneous to my will. This is just a description of heteronomy: to allow something other than one's will to determine one's decisions. Therefore, to *treat* your reasons as reasons – to *treat* you as an end in yourself – is an instance of heteronomy.

Conversely, for my decision to do what you say to be based on the application of the categorical imperative, is for me to determine whether the maxim of my action is a fit principle. That is something that I decide in full expression of my normative status for myself. In doing so, I am refraining from treating you as an end for me. It seems that the expression of my autonomy requires that I don't treat you as an end for me.[13]

One might point out that the way I presented the example of autonomously deciding to eat the chocolate obscures some details which might affect my conclusion that I cannot autonomously treat you as an end. These details concern what goes into the decision to eat the chocolate. In my presentation of the example I said that my decision to eat the chocolate is autonomous only when I conclude that the maxim which would capture my action passes the categorical imperative test. However, that a given maxim passes the categorical imperative test makes that decision permissible, but not necessarily required.

Presumably, the principle behind my decision to eat the chocolate falls into the category of maxims which are fit as principles, but that are not required. Let's assume that that is the case. In that case, my application of the categorical imperative assures me that, were I to eat the chocolate (on the given maxim), I would be expressing my autonomy. But there is a gap between deciding that eating the chocolate is ok, and my decision to eat the chocolate. If I make the decision to go ahead and eat the chocolate, what bridges the gap between the decision that it is ok to eat the chocolate and the decision to do it? Why do I make that decision? I make that decision because I have concluded that doing so is permissible *and* because I fancy having the chocolate. Two motives come into my decision: one my desire for the chocolate cake, and another the conclusion that deciding to have the chocolate cake is permissible. But we still think that this decision expresses my autonomy.

[13] This bears directly on the conclusion tentatively reached in Section 4.4.1. that one is required to value others in as long as one values them. I address this point below, in Section 5.5.1.

In the same vein, we might think that my decision to do what you say might be the result of both my conclusion that doing what you say to do is permissible *and* the claim on my will to do as you say. I conclude that parking my car down the road, instead of outside your door, is permissible *and* there is a claim on my will to not park the car down the road. So I decide to part my car down the road.[14]

This duality of motives is analogous to the duality involved in my decision to eat the chocolate. If the participation of the application of the CI in the formation of my decision to eat the chocolate is sufficient for that decision to be expressive of my autonomy, then the participation of the application of the CI in the formation of my decision to park my car down the road will also render my decision expressive of my autonomy. But if my autonomous decision to park my car down the road also includes my seeing you as an end, then it seems that that decision is both autonomous and treats you as an end.

However, that decision does not treat you as an end. That is because my seeing you as an end - i.e. the claim on my will to do as you say to do just because you say it - plays only a contingent role in the formation of my decision. That is, my seeing you as an end does not determine my decision to do what you say to do. But, as I explained in p.98 above, and in Section 2.2.1., for me to see someone as an end is for me to see them as placing unconditional constraints on my choice of actions - as setting ends for me. In the current example, my seeing you as an end has not set unconditional constraints on my choice of actions. I would have not chosen to park my car down the road had I not decided, *independently of your normative status to me*, that doing so is ok. So, in this example, whilst I *see* you as an end, I have not *treated* you as an end.[15]

This point might better meet our intuitions in the following example. Suppose I'm a beggar in the street. You see me as you walk past me and I place a claim on your will to help me. You are too attached to your money and would rather pretend that you haven't seen me, or that you don't feel a 'pull' to help me - i.e., that I haven't placed a claim on your will. However, you have this loose change in your pocket which is rubbing uncomfortably as you walk. You decide that it would be a good thing (i.e. permissible) to get rid of that change by

[14] Herman uses this structural point to defend the possibility of a Kantian engaging in genuine friendship (Herman 1983).

[15] Alternatively, we could say that, in fact, I do not *see* you as an end. This would yield different implications for Korsgaard's attempt to establish morality. However, since in order to *treat* you as an end I have to *see* you as an end, we would end with the same conclusion I am arguing for.

giving it to me. And since I have placed a claim on your will to help me, you do hurl the change at me.

I take it that we wouldn't say that you have treated me as an end. My argument would vindicate that intuition. Although your seeing me as an end has come into your decision to give me your change, its role in your decision is subordinate to your decision that getting rid of the change would be a good thing. You would not have given me the money had you not decided that doing so would rid you of discomfort. But in that case you have not treated the claim I make on your will as setting unconditional constraints on your choice of actions. Your giving me the money has been *ex hypothesis* conditional on it being ok to rid yourself of the burden of carrying it. So you have not treated me as an end.

For me to autonomously treat you as an end I would have to decide to do as you say because I had concluded that you are an end, and that you are to be treated as such. However, as I explained in Section 3.7.1., it follows from the constructivist take on normativity as a claim on the will that I can't tell whether you are a source of normativity unless you make a claim on my will. And, as we saw in Section 2.3.1., the idea that reasoning can lead me to having a claim on my will by you is confused. So reasoning can't lead me to see you as an end. Therefore, if I *conclude* that you are an end, I must have reached that conclusion through bad reasoning. If I reach that conclusion through bad reasoning, that reasoning does not qualify as expressive of my autonomy, as my autonomy is expressed through correct reasoning. So the corresponding decision would not be an expression of my autonomy.

Nor would it be an instance of my treating you as an end. To treat you as an end is to abide by the normative status you have for me - by your status to me as an end. But I cannot reason my way to seeing you as an end. So, if I concluded that you are an end for me, not only would I have performed faulty reasoning, but I wouldn't have got myself to seeing you as an end either. If I don't see you as an end, I cannot treat you as an end. So my decision to treat you as an end because I have concluded that you are an end and to be treated as such, would be neither an expression of my autonomy, nor a way of treating you as an end.

If my arguments are correct, to treat you as an end constitutes an instance of heteronomy because it consists in allowing my decision to be determined by something extraneous to my will. However we might think that we can challenge this conclusion by appeal to a

qualification that Korsgaard makes to the idea that to treat you as an end is to do what you say just because you say it. (This is the important condition I mentioned above (p. 112).)

Korsgaard explains that we might treat others as ends even if we don't do as their reasons say to do. You call me, but I'm running for the bus, so instead of coming to you I yell that I'll call you later (SN 140). You ask me to not park my car outside your door, but I explain to you that all parking spaces in the street are outside someone's front door, and barring some special reason, you can't expect me not to park outside your door any more than the rest of us can expect anyone else not to park outside our door.

According to Korsgaard, these examples still count as my treating you as an end: although I don't do what your reasons say to do, I instead offer you my own reasons why not (SN 140). I would only do that as a way of acknowledging to you your normative status for me. My aim would be to let you know that I'm not doing as your reasons say to do not because I resist treating you as an end, but rather because something prevents me from doing the specific actions your reasons say to do. But for me to respond to your normative claim by acknowledging to you the normative status you have for me is for me to treat you as an end.[16]

This might seem to challenge the conclusion that to treat you as an end is an instance of heteronomy. That conclusion was based on the idea that in treating you as an end I allow my decisions to be determined by something extraneous to my will, namely, the claim you make on my will. Yet we have now seen that I might decide whether to do as per your reasons or not. We might think that if I decide to give you a reason why I won't do as your reasons say to do, that decision cannot have been determined by your claim on my will. Since, according to Korsgaard, to give you a reason why I won't do as your reasons say to do is to treat you as an end, if I my decision to give you such a reason has not been determined by your claim on my will, then it is at least possible that I might have reached that decision in a way that expresses my autonomy.

However, it is incorrect to think that my decision to give you a reason why I don't do as your reasons say to do has not been determined by your claim. This is because the selection of possible actions over which I deliberate in those cases is constrained by the requirement that they be ways of treating you as an end. So I am allowing your claim, and not the

[16] This discussion is developed over SN pp.140-142.

categorical imperative, to determine the range of actions that I might choose to perform. Furthermore, if treating you as an end is incompatible with the expression of my autonomy, as I have argued, then to deliberate about the way in which to treat you as an end cannot be expressive of my autonomy either.

We find an analogy of this last point in desires. Suppose I have decided rightly that I ought to not drink anymore and go home instead. Yet, despite my decision, when I am next asked whether I would like some more red or white, I give in to temptation and, after some deliberation, I decide for the white. I deliberate alright, but the options I have deliberating over would both instantiate heteronomy, just slightly different from each other – in the one case, I am violate my decision to go home by having some more red; in the other case by having some more white. In either case my decision is determined by my desires, and thus I am being heteronomous.

So the idea that we might not do as another's reasons say to do but treat the other as an end nevertheless, does not contradict the conclusion that to treat another as an end is to act heteronomously.

5.4.1. Public reasons and the moral law

Kant thought that the moral law is the law of reason, which is the categorical imperative. He also thought that the agent's autonomy is expressed through the application of the categorical imperative. So, according to Kant, to express one's autonomy is to apply the moral law. However, due to the conception of practical reason she espouses, Korsgaard thinks that the law of reason and the moral law are not the same. They are both instances of the categorical imperative, but the difference between them is the scope of application of the categorical imperative. The scope of application of the law of reason is the agent alone, whilst the scope of application of the moral law is all agents. So Korsgaard's revision of the categorical imperative disturbs the connection with autonomy that was part of Kant's thought. Contrary to Kant's thinking, the expression of one's autonomy does not necessarily consist in the application of the moral law.

Korsgaard's own efforts to establish a constructivist morality would, if successful, bring together the agent's autonomy and the moral law. The agent's autonomy and the

categorical imperative are connected in that the agent's autonomy comprises the sources of normativity for the agent, and the scope of application of the categorical imperative comprises the sources of normativity for the agent too.

That both the agent's autonomy and that the scope of application of the categorical imperative comprise the sources of normativity for the agent is not a coincidence. Since the agent expresses her autonomy though the application of the categorical imperative, the scope of application of the categorical imperative is determined by the range of sources of normativity comprised in the agent's autonomy. In this way, the law of reason applies over the agent herself alone because according to Korsgaard's account of reason the agent *need* see only herself as a source of normativity. That is, the expression of her autonomy requires that the categorical imperative be applied only over herself.

Since the moral law applies over all agents, including oneself, if Korsgaard is going to connect it to autonomy, she will have to show that the agent's autonomy comprises not just herself as a source of normativity for herself, but all agents. For the agent's autonomy to comprise not just herself as a source of normativity, but all other agents, it will have to be the case that the agent sees not just herself but all other agents as sources of normativity for her.

That is just what the account of the publicity of reasons is supposed to show. In arguing that we are susceptible to each other's reasons, Korsgaard argues that we are normative to each other. And this gives rise to the moral identity, where the agent conceives of herself and of all other agents as normative for herself. In other words, the moral identity comprises the agent's autonomy when the agent's autonomy has all agents as sources of normativity for the agent. As such, the moral identity would require that the categorical imperative be applied to all agents. That is, from the moral identity we get the categorical imperative as the moral law. So, by establishing the moral identity, Korsgaard would have brought back Kant's connection between autonomy and the moral law. It would be a revised connection – she retains a conceptual difference between the moral law and the law of reasons, as well as between the agent's autonomy as comprised in the human identity and the agent's autonomy as comprised in the moral identity – but it would be sufficient to maintain Kant's core idea that the moral law springs from our humanity – from our status as ends in ourselves.

In Section 4.4.1. I argued that that objective was slighted thwarted by the fact that Korsgaard does not show that necessarily all agents have a moral identity.[17] But in the section above I have argued that treating others as ends is an instantiation of heteronomy. And this inserts an ineliminable division between autonomy and the moral law. This is because to act heteronomously is to fail to express one's autonomy. And to express one's autonomy is to treat oneself as an end. If one cannot at the same time treat others and oneself as ends, then one cannot express one's autonomy by treating others as ends.

But that is what the moral law requires us to do. The moral law requires that the universalization test applies to all agents, including oneself, and to apply the universalization over all agents is to treat them as ends. So the moral law requires that the agent treat all agents, including herself, as ends. But if treating others as ends prevents the expression of the agent's autonomy, it follows that the agent cannot express her autonomy through the application of the moral law. So Korsgaard's account has decisively severed Kant's connection between autonomy and the moral law.

More than that, if it is the case that treating others as ends is incompatible with treating oneself as an end, since that is what the moral law requires that we do, it follows that, under Korsgaard's account of reason, the moral law is incongruous. This has far reaching implications. One of them is of particular interest to us, as it concerns something we have encountered at several points throughout this thesis so far. This is the argument for the Formula of Humanity.

The Formula of Humanity is the formulation of the categorical imperative that most explicitly says to treat others and oneself always as ends. I rehearsed the problems critics have found with the argument intended to support that formulation in section 2.3.1. However, if the idea of treating others as ends is irreconcilable with the idea of treating oneself as an end, then the problem with the Formula of Humanity is not so much that it doesn't have an argument to support it, but that it is incoherent. So, Korsgaard's separation between the categorical imperative as the moral law and the categorical imperative as the law of reason both disturb Kant's connection between autonomy and the moral law, and challenges the very idea of the moral law.

[17] I will return to this point as it concerns Korsgaard's own account in the next section.

5.5.1. Public reasons and the human identity

The conclusion that, given the thesis of the publicity of reasons, to treat others as ends is to instantiate heteronomy has implications for Korsgaard's own account too. In Chapter 4 I argued that even if Korsgaard has not established that all agents have a moral identity, her account has the resources to give us the conclusion that very nearly all agents *do* have a moral identity. From there I offered a construal of the relation between valuing others and valuing oneself such that, if one values others one is required to value them. If this is so, since one's valuing of others is comprised in one's practical identity – the moral identity – it follows that one ought to have a moral identity *if* one has a moral identity. Since we had previously concluded that very nearly everybody has a moral identity, then that yields the conclusion that very nearly everybody is required to have a moral identity.

However, if it is the case that treating others as ends is an instance of heteronomy, that conclusion is challenged in a different way. That is because, if it is the case that treating others as end is an instance of heteronomy, one is required to *not* have a *practical* moral identity. Let me explain.

The human identity comprises the agent's own valuing of herself as an end, so it involves the agent's seeing herself as an end. Seeing herself as an end is what the agent's autonomy consists in. Therefore the human identity comprises the agent's autonomy. If the human identity comprises the agent's autonomy, then to express the human identity necessarily involves the expression of the agent's autonomy. If treating others' reasons as reasons – treating others as ends – goes against one's autonomy, it goes against one's human identity. So one's human identity *requires* that one does not treat others as ends.

Now, as for the moral identity, I have so far been talking about the moral identity as a practical identity. A practical identity, we will recall, is a description which one applies to oneself *and* which one sees as normative for oneself.[18] This allows for self-conceptions, for descriptions of oneself, which are not normative. I have brown eyes, I prefer coffee to tea, and I was born in August. All those things are true of me, but none of them form a practical

[18] This has realist overtones, but to say that one finds a certain self-conception, or self-description, normative for oneself is just to say that one is normative for oneself under a certain description which one values. The source of value keeps coming back to the agent. This was explained in Chapter 3, and remains an implicit assumption throughout the thesis.

identity for me – I don't look to those descriptions of myself as values through which to guide my actions.[19]

The moral identity, however, is supposed to be *practical* because the description of myself contained there is normative to me. I do look to that description as a value through which to guide my actions. However, an implication of the conclusion that treating others as ends instantiates heteronomy is that one is required to not treat one's moral identity as a practical identity. That is because the moral identity consists in the agent seeing both herself and others as ends in themselves.

As we saw above, whether I see others' reasons as reasons for me is not something I decide. So whether the moral identity *as a description* applies to me is not something that I can decide. What I can decide is whether that identity is *practical* – whether it is a description of me which I take to be normative and through which I guide my actions. But to guide my actions by my moral identity involves treating others as ends, and thus going against my own human identity. So my human identity requires that I not treat my moral identity as normative.

5.6.1. Conclusion

I conclude that Korsgaard's attempt at establishing morality has not been successful. The main tenet of constructivism is that normativity originates in the agent, and that tenet presents a challenge for how to incorporate the normative status of others into one's normative landscape. Korsgaard attempts to do this by arguing that we are subject to the normativity of others just as we are to our own. However, Korsgaard's account of normativity puts the other's normative status on a different footing from one's own. The special status of one's own normative status for oneself is due to its connection to the agent herself. But this connection is necessarily absent from the claim which another's normative status places on my will. That makes the claims placed on the agent by the normative status of others subject to scrutiny just like the claims placed on the agent by anything other than her own will, for example, by desires. But I have argued that to submit to scrutiny the claims that the normative status of others places on one's will, is to disregard the others'

[19] This was explained in more detailed in Section 3.4.1.

normative status. Correspondingly, to upheld the normative status of others is to disregard one's own normative status to oneself. In other words, to upheld the normative status of others is to be heteronomous. Therefore, one's autonomy requires that one not upheld the normative status of others.

Stephen Darwall develops an account of morality which is in some important ways similar to Korsgaard's. He too seeks to base morality on a normative relation between agents. What is of special interest to me is that one of the purported strengths of his account is that it provides the way in which the agent's autonomy is uniquely realised. Darwall's account is not a constructivist one,[20] but it is affine to it in some respects. If his account were successful, we might be able to transpose it to Korsgaard's constructivism, and thus obtain a constructivist morality. I spend the next chapter looking at whether Darwall is successful in that aim.

[20] It might be seen as a constructivist account itself, but as I shall point out one of the main premises of his argument stands in direct contradiction to constructivism.

- 6 -

Darwall's second-person standpoint and autonomy

6.1.1. Introduction

I reached the end of the previous chapter on the conclusion that Korsgaard's account of the publicity of reasons advances an account of morality in conflict with the idea that one is to be normative to oneself. If I treat your reasons as reasons for me, I argued, I must sidestep the normative status I have for myself, i.e. my autonomy. If, on the other hand, I express my autonomy, I must sidestep any normative status you or your reasons have for me. As such, my autonomy requires that I don't treat you as an end. This conclusion violates the connection in Kant's thought between autonomy and the moral law. But it also shows that Korsgaard has not been able to establish a constructivist morality because it violates the main constructivist tenet that the agent is to be normative to herself.

With that in mind, Stephen Darwall's account of the second-person standpoint ('SPS' henceforth) is a natural place to turn to next. Although Darwall does not endorse constructivism, his account of the SPS is in many ways close to Korsgaard's account of the publicity of reasons.[1] Darwall's SPS, like Korsgaard's public reasons, attempts to develop an account of the way in which agents are normative to each other - of agents seeing and treating each other as ends. Also like Korsgaard, Darwall does this as a way of addressing what he sees as a deficiency, or underdevelopment, in Kant's moral theory.

What makes Darwall's account particularly attractive for my project is that he argues that the expression of the agent's autonomy is uniquely facilitated by the SPS. Given the shared ground between Darwall's and Korsgaard's accounts, we might be able to transfer the

[1] Darwall does however briefly touch on the thought that some aspects of his account would find a natural metaethical home within constructivism (SPS 291-297). But he doesn't develop or defend constructivism *per se*.

relevant aspects of Darwall's account to Korsgaard's and thus gain some ground towards a constructivist morality. My aim in this chapter is not so much to examine the virtues of Darwall's account as it is. Rather, my aim is to see whether it can help the constructivist (at least the Korsgaardian constructivist). My conclusion will be that it can't.

I have said that Darwall's account attempts to improve on Kant's moral theory. I begin by presenting the problem which Darwall sees with Kant's moral thought, so we can better appreciate the solution he proposes.

6.2.1. Morality and reason

Here is a thumbnail sketch of Darwall's overall argument. According to Kantian moral theory, the requirements of morality are requirements of reason.[2] If it is the case that the requirements of morality are requirements of reason, then any requirement of morality will be fully supported by reason alone. But, according to Darwall, we can't fully support the requirements of morality by reason alone. Therefore, it cannot be the case that the requirements of morality are requirements of reason alone. That is Darwall's negative account. His positive account is that reason supports the requirements of morality only within the SPS. I shall now develop the main parts of Darwall's dialectic.

Darwall thinks that reason alone cannot support the requirements of morality on two counts: reason cannot give us the idea that persons are ends in themselves; and reason cannot supply the bindingness, or normativity, of the moral law (S. L. Darwall 2006, pp.138–139)[3]. Those requirements of morality are contained in Kant's Formula of Humanity: 'So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means' (GW 4:429). So, we might put Darwall's worry that reason cannot give us either the content of the Formula of Humanity, nor its connection to the will – its normativity. But Kant and many Kantians think that reason can give us both. So let us look at why they think that, and where Darwall thinks they err.

[2] Unless otherwise indicated, by 'reason' and its cognates, I mean 'practical reason' and its cognates, as appropriate.

[3] Subsequent references to Darwall's (2006) *The Second-Person Standpoint*, will be given with the abbreviation SPS.

Kantians argue that reason grounds both the content and the normativity of the Formula of Humanity because they both arise from inescapable presuppositions of the practical standpoint. That is, they argue that one could not be engaged in reasoning unless one adopted presuppositions which led directly both to the normativity of the Formula of Humanity, and to its content of treating persons as ends in themselves. I will first look at how the presuppositions of the practical standpoint are supposed to lead to the normativity of the Formula of Humanity - to its connection to the agent's will. After that, I will look at how those presuppositions are supposed to lead to the content of the Formula of Humanity - to the idea that persons are ends in themselves.

6.2.2. The normativity of the moral law

The presupposition which is supposed to lead to the normativity of the Formula of Humanity is the presupposition of autonomy. This is because, according to Kantian thought, the moral law (of which the Formula of Humanity is a formulation) is normative for the agent because it is the law of the agent's own will.[4] On this view, for the agent to apply the law of her will is for the agent to express her autonomy. So, for the agent to express her autonomy is for the agent to apply the moral law, and for the agent to apply the moral law is for her to express her autonomy. This is what is termed 'Kant's Reciprocity Thesis', the thesis that autonomy and the moral law entail each other.[5] If autonomy and the moral law imply each other, then if there is such a thing as autonomy, there will be such a thing as the moral law, and vice versa. Having set out this thesis, Kant goes on to show that there is such a thing as the moral law by arguing that there is such a thing as autonomy.[6]

Kant argues that there is such a thing as autonomy by arguing that autonomy is a necessary presupposition of the practical standpoint. We have already come across Kant's argument for autonomy (Section 5.2.1.), so I will be brief. The practical standpoint necessarily presupposes autonomy, because, Kant maintains, it is part of what it is to engage in

[4] As we know from Section 3.4.3. Korsgaard challenges the Kantian idea that the law of the will - the law of reason - is the moral law.

[5] I believe the term was coined by Allison (1986).

[6] He does it this way in GW. In the *Critique of Practical Reason* he proceeds to argue the other way round. Darwall discusses both, but I will limit myself to his engagement with the GW, as that is sufficient for the dialectical shape of this thesis.

deliberation that the agent presupposes that it will be *herself, through her deliberation* that will determine the conclusion of her deliberation. If the conclusion of the agent's deliberation is normative, and the agent presupposes that that conclusion is reached by the agent herself alone, then the agent presupposes that she is normative - the agent gives herself her own instructions, or laws. Hence, the agent presupposes that she is autonomous. She presupposes that she governs her actions.

According to Kant, decisions are reached on the basis of principles. So, as part of presupposing that she will determine the conclusion of her deliberation through her deliberation, the agent presupposes that the principle that will produce her conclusion will be a principle of reason. The principle of reason is the categorical imperative. So, as part of presupposing that she will determine the conclusion of her deliberation through her deliberation alone, the agent presupposes the application of the categorical imperative. So, in presupposing her own autonomy, the agent presupposes the application of the categorical imperative.

Kant talks about these aspects of autonomy as freedoms: the freedom to reach one's own conclusion through deliberation, and the freedom to apply the will's own law to one's deliberation.

If Kant's arguments were successful in establishing autonomy, the moral law will also be established, and with it, the normativity of the Formula of Humanity. But according to Darwall the freedoms which Kant shows are presupposed in deliberation do not amount to autonomy (SPS Chapter 9). So, according to Darwall, Kant's argument does not establish that autonomy is grounded in the presuppositions of practical reason. And, as a corollary, he doesn't show that the normativity of the moral law is grounded on the presuppositions of practical reason.

Darwall thinks that Kant is right that the practical standpoint presupposes the freedom from causal interference with the agent's reasoning, and the freedom to deliberate in accordance with principles of reason (SPS 214-215). But Darwall does not think that these freedoms amount to autonomy. For those freedoms to amount to autonomy, they would have to lead the agent directly to the idea that her will is her source of normativity - that what reasons she has are ultimately determined by the law of her will. Since Kant equated the will with practical reason, we might also say that those freedoms would amount to

autonomy only if they led the agent to the idea that her reasons receive their normativity from her own reason. But Darwall does not think that they do.

Darwall does not think that those freedoms necessarily lead the agent to the idea that she is the source of normativity of her reasons - her own autonomy - because of the possibility of what he calls the 'naïve reasoner'. The naïve reasoner is someone who displays the freedoms highlighted by Kant, whilst thinking that her reasons are sourced outside her will. The naïve reasoned displays the freedoms highlighted by Kant in that, when engaging in deliberation she thinks that she will reach the answer to the question of what to do by the employment of her reason: she doesn't think that the answer to the question of what to do will be reached by tossing a coin, for example. And she also thinks that she is a competent reasoner. In other words, the naïve reasoner thinks that she can reach the answer to the question of what to do by herself.

However, the naïve reasoner does not presuppose that her will is the source of her reasons. Instead she thinks that her reasons are grounded on the objects of her volition. It is not that the naïve reasoner thinks that her desires are her reasons. Rather, she takes her desires to track what reasons there are for her (SPS 225). For example, if she has a desire to go to the cinema tonight, she takes it that she has a reason to go to the cinema tonight. That is because she thinks that her desire to go to the cinema tonight indicates that there is a reason for her to go to the cinema tonight (SPS 225).

Darwall's naïve reasoner, then, is a counterexample to Kant's argument for autonomy. The naïve reasoner presupposes the freedoms which Kant thought yield the idea of autonomy, whilst thinking that her reasons are sourced outside her. Since the idea of autonomy is the idea that one is one's own normative source, it follows that the naïve reasoner does not think of herself as autonomous. It doesn't matter if the naïve reasoner is wrong in her assumption about the source of her reasons. What matters is that so long as we think of her as a coherent reasoner, we are committed to thinking that the freedoms presupposed in deliberation do not inescapably lead to the assumption of autonomy (SPS 226-227).

As I mentioned above (on page 125), since Kant thought that the moral law and autonomy entail each other, he attempted to establish the moral law by establishing autonomy. If he showed that autonomy is an inescapable presupposition of the practical standpoint, he will have shown that the normativity of the moral law is an inescapable presupposition of the practical standpoint. But according to Darwall the possibility of the naïve reasoner shows

that it is not the case that autonomy is an inescapable presupposition of the practical standpoint. If autonomy is not an inescapable presupposition of the practical standpoint, the moral law isn't either. And, if the moral law is not an inescapable presupposition of the practical standpoint, we cannot account for its bindingness by appeal to the practical standpoint.

This is why Darwall thinks that reason is deficient to yield the bindingness, or normativity, of the moral law. But, as I mentioned above, Darwall also thinks that reason cannot deliver the content of the moral law either.

6.2.3. The content of the moral law

According to Darwall, the support for the view that reason gives us the idea of treating agents as ends relies on the view that practical standpoint presupposes autonomy. Since, as we have just seen, Darwall argues that the practical standpoint does not presuppose autonomy, he also thinks that the view that reason gives us the idea of treating agents as ends is unsupported. He concludes that reason cannot supply the idea of treating agents as ends.

To show that the support for the view that reason gives us the idea of treating agents as ends relies on the view that practical standpoint presupposes autonomy, Darwall uses Korsgaard's reconstruction of Kant's argument. That argument is as follows.

When we make a choice, we take the object of our choice to be valuable. When we take the object of our choice to be valuable, we understand that if it weren't for our desires and inclinations we would not find those objects valuable. So, we think that the object of our choice is valuable only because we value it. If we think that things are valuable only because we value them, we must 1) think that we are sources of value; and 2) value ourselves whenever we value anything. If things are valuable because we value them, and we value ourselves, then we are valuable. And if we are the sources of value, then our value is not instrumental, but final - we are ends in ourselves. Kant adds that 'The human being

necessarily represents his own existence in this way [i.e. as an end in itself]', and Korsgaard concludes that 'in this way, the value of humanity itself is implicit in every human choice'.[7]

As Ridge has pointed out (Ridge 2005b) that is a controversial argument. However, my aim here is not to examine whether that argument is defensible or not. Rather, my aim is to show what Darwall thinks is wrong with it so we can understand better his attempt to fix it, viz. his account of the SPS.

As Darwall points out, Korsgaard's/Kant's argument assumes that the practical standpoint presupposes autonomy in its rejection of the idea that the agent might think that the objects of her choice might be valuable independently of being valued by herself (Darwall 2009, p.149). But we have just seen that Darwall argues that the practical standpoint does not necessarily presuppose that the agent herself is the source of value of her objects of choice. The naïve reasoner presents a picture in which an intelligible reasoner takes the objects of her choice to be valuable independently of her. So, according to Darwall, Korsgaard/Kant can't help themselves to the premise that as part of engaging in deliberation one must think that one's objects of choice are valuable because one values them. That is, Korsgaard/Kant can't appeal to the idea that in valuing something S the agent sees herself as the source of value of S. Without that premise, they cannot derive the idea that one takes oneself to be valuable – that one takes oneself to be an end in oneself.[8] And, without that idea, Korsgaard/Kant cannot derive the idea that we are to treat agents as sources of value, i.e. as ends in themselves.

This, then, is why Darwall thinks that reason cannot deliver either the normativity of the moral law, or its content. To recap, Darwall agrees that the practical standpoint does assume some freedoms, but he argues that those freedoms do not amount to autonomy. Since autonomy and the moral law imply each other, if the presuppositions of the practical standpoint do not presuppose autonomy, it follows that they do not presuppose the normativity of the moral law. If the presuppositions of the practical standpoint do not

[7] Korsgaard's argument appears in SN, p.122, and is quoted by Darwall in SPS, p. 230, and in 2009, p. 149. Kant's quote is from GW 4:428-429. As we know from Chapter 2, Korsgaard does not believe that that argument establishes that reason yields the idea that the agent is to treat *all* agents as ends, but only that she is to treat herself as an end. To show that reason yields the idea that all agents are to be treated as ends, she develops her thesis of the publicity of reasons. Darwall does not note this aspect of Korsgaard's account. I won't dwell on that fact, as it is immaterial for my purposes.

[8] In a footnote, Darwall sides with critics who object to the idea that if something is a source of value it is therefore valuable in itself (SPS 231n30). He doesn't pursue that objection, though, and nor will I. However, I will mention it again below.

include the normativity of the moral law, we are not bound to the moral law through the presuppositions of the practical standpoint.

The content of the moral law is presupposed by the practical standpoint only if autonomy is presupposed from the practical standpoint. That is, the agent thinks that humanity is an end in itself only if she thinks that she is the source of her reasons - i.e. if she is a source of normativity. Since the agent is not necessarily committed to the thought that she is a source of normativity, she doesn't necessarily obtain the content of the moral law.

If reason can't deliver either the normativity of the moral law, or its content, we will have to look elsewhere. The place to look is, according to Darwall, a particular standpoint – what he calls 'the second-person standpoint'. I present Darwall's account of the SPS below.

6.3.1. The second-person standpoint

The main characteristic of the SPS is that in it agents make, and are subject to, a certain kind of direct claim on each others' wills (SPS 3, 5). Darwall invites us to recognise this type of direct claim through the following example. Suppose you are stepping on my foot and I ask you to stop doing so. I might intend for my request to work in one of several ways. I might intend for my address to point out to you that you are singularly well placed to correct a fact about the world which is not as it should be (viz. your foot is on top of mine and it shouldn't be) (SPS 5-7). Darwall suggests that we see the role of my address here as epistemic: my aim is not so much to *give* you a reason, as to point out one that is there anyway. Viewed this way, by asking you to remove your foot from the top of mine, I aim to give you an epistemic reason (SPS 6-7). As such, that reason would be valid for anyone who was in the position of bringing it about that my foot is not stepped upon.[9]

I could, alternatively, aim to charm you, or to intimidate you into removing your foot from the top of mine.[10] Here, my aim would be to generate a situation relevant to your interests, where pursuing your interests would serve mine. If you want my charm, or if you want to

[9] We don't have to think that the implicit conception of reasons at play here is ultimately defensible. The point of this example is to provide a contrast to the distinctiveness of second-personal reasons.

[10] This is suggested in SPS 39-43.

avoid my ire, you had better remove your foot from the top of mine. We could say that here my aim is to give you a prudential reason.

As a third alternative, I could aim for my address to be a piece of advice. Perhaps I know that you have a good heart, and that it would upset you to inadvertently step on another's foot. Here I would be aiming to convey to you that it might be a good idea for you to remove your foot from the top of mine (SPS 49). None of these ways of addressing my request to you take place within the SPS, according to Darwall.

My address is a second-personal address - it aims to make a direct claim on your will - when I intend for my address to get you to remove your foot from the top of mine because I am asking you to *and* you recognise my *de jure* authority[11] to ask you to do this. Specifically, I intend for you to recognise that my address is authoritative in virtue of my being a person, or an agent, or a member of the moral community, or in virtue of satisfying some such normative loaded description; and that it is authoritative to you in virtue of your also being a person, or an agent, or a member of the moral community, of in virtue of satisfying some other normative description.[12] In other words, I expect you to see me and to see yourself as having an authoritative status to each other, and I expect you to take my address as an expression of my authority (SPS 7-9).

Notice that the insistence that you recognise my (*de jure*) authority is important, as it differentiates the second-personal address from improper uses of authority, such as coercion. If you do what I say because of fear of, or desire for, the consequences, you are guiding your actions by their expected outcomes. If you do what I say because you recognise my authority, you recognise that I am using my authority properly. To use my authority properly would be to address your own authority to self-impose my claim just in virtue of your recognising my authority (SPS 21). This is what Darwall calls 'Fichte's Point', and it is what ensures your self-determination by ensuring that you self-impose my second-personal address *freely* (SPS 20). If you didn't self-impose my address freely, you would not hold yourself into account should you fail to do as per my address (SPS 258).

Notice too that the reason given here is agent-relative (SPS 8). Even if someone else were able to correct the state of affairs of my foot being stepped on by yours, it is *you* I am asking

[11] Unless otherwise indicated, I mean any mention of authority *de jure*.

[12] Throughout the text, and unless otherwise stated, I shall use all terms referring to persons interchangeably and as carrying normative import.

to do it, and *your* recognition of my authority that I am summoning. A direct claim on another's will, viz. the distinctive second-personal claim, is one which does not obtain its normative input from its relation to epistemic reasons, or to prudential concerns, but from the relation between the issuer of the claim and the receiver of the claim. This relationship is one of mutual recognition of the authoritative status of each. The claim I aim to make on you bypasses all the processes involved in the alternative ways of address listed above. Instead it relies on your recognising my authority and thus imposing my claim onto you freely.

In light of this characterisation, what Darwall means by my *making a direct claim on your will* is my giving you an end, a reason, backed by my normative status for you. If we put it that way, for me to make a direct claim on your will, is for you to see me as having a normative status for you, which is to see me as an end in myself. Since the main characteristic of the SPS is that we make and receive direct claims on each others' will, we might also put it that the main characteristic of the SPS is that in it we are normative for one another, we have a status of ends in ourselves for each other. So, Darwall's solution to the problem that the practical point of view cannot deliver the idea of persons as ends in themselves, is to say that seeing ourselves as ends in ourselves for each other is just a *way* of seeing ourselves and each other, a standpoint, and not a conclusion delivered by practical reasoning.[13]

From this standpoint emerge all the rules and features constitutive of morality; and our moral practices, such as seeking accountability or atonement as well as directing Strawsonian reactive attitudes towards each other, are explained and justified. Since I see myself as an end for you, I expect you to take my address as an expression of my authority to you. And since I expect you to take my address an expression of my authority to you, I expect you to take yourself to be accountable to me (and to you) in respect to my address. That is, I expect you to, as a person, freely impose upon yourself the normative claim that I, as a person, make on you (SPS 248). If you recognise my authority but not respect it – if you *see* me as an end, but fail to *treat* me as an end – I shall seek to hold you to account. That is, if you don't do as per my request, I shall insist that you recognise my authority towards you by doing something else, typically, offering me an explanation, seeking my forgiveness, or offering reparation of any damage caused, including any damage caused to our relationship

[13] This is one of the points in which Darwall and Korsgaard are in agreement. Exploring the similarities and differences between the two accounts would be a rich exercise, but it falls out of my remit here.

as ends for each other.[14] Also, I am resentful towards you, thus implying that I think that you have not treated me as en end in myself – that you have not abided by my authority to you (SPS 15).

If the SPS is characterised by agents seeing each other as ends for themselves and for each other, we can tease out the following presuppositions implicit in the SPS. Once we see how the rules and features constitutive of morality and of our moral practices arise from that standpoint, we obtain the following set of presuppositions. The agent presupposes that both she and the other are free and rational agents; that she and the other have authority over each other; that they each can accept this authority; and that they each are responsible for responding appropriately to each other's authority.

These presuppositions are all tightly related. To have authority over someone is to have a normative status over that person. Given any construal of normativity, to be subject to a normative claim to ϕ is to be under an obligation to ϕ. And to be under an obligation to ϕ is to be accountable for ϕing. If I am the source of your normative claim to ϕ, then you are accountable to me. If you are the source of that normative claim, then you are accountable to yourself. Since, according to Darwall, what makes me normative to you and what makes you responsive to my normative status is both your and my freedom and rationality, you respond to my normative status freely and rationally - that is, you freely impose my normative claim upon yourself. So, you and I are both the source of that normative claim. Therefore, you are accountable to both you and me. A second-personal address is successful when those presuppositions are correct (SPS 11-15).

Darwall develops the SPS to account both for the normativity and for the content of the moral law as formulated in the Formula of Humanity. From what we have seen so far it is quite obvious how the SPS grounds the content of the moral law: seeing ourselves as ends in themselves to each other is a necessary presupposition of the SPS. My main interest, though, is how the SPS is supposed to ground the normativity of the moral law - how it is supposed to uniquely realise the agent's autonomy. To that I turn now.

[14] This is the same point that Korsgaard makes when she said that in offering you a reason why I can't/don't do as your reasons say to do, I am treating you as an end (Section 3.7.1.).

6.4.1. The SPS and autonomy

Darwall thinks that the SPS realises the agent's autonomy because the agent's autonomy is a necessary presupposition of the SPS. Like all other presuppositions of the SPS, it derives from its main characteristic, namely, that we are normative to one another.

We saw that Darwall argues that practical reason does not commit us to autonomy because it is intelligible for a non-second-personal deliberating agent to assume that her reasons are grounded on the objects of her volition. But within the SPS, that possibility does not obtain.

For you to occupy the SPS is for you to see yourself and others as normative to each other. For you to see another as having a normative status for you is for you to see yourself under the normative claim from the other regardless of whether it suits your interests. If for you to see another as having a normative status for you is for you to see yourself under the normative claim of the other regardless of whether it suits your interests, then seeing another as having a normative status for you rules out your thinking that your reasons are grounded in the objects of your volitions.

In other words, in recognising a second-person address, you must accept that you are capable of acting on it regardless of your desires (Darwall 2007, p.58). So, unlike the non-second-personal reasoner, who can intelligibly think that her reasons are grounded in the objects of her desires, a second-personal reasoner cannot intelligibly think that her reasons are grounded in the objects of her desires. And, since it is constitutive of responding second-personally to an address that the agent recognises the authority of that address, the agent imposes that address upon herself *freely* (SPS 258-259). If the agent did not impose that address upon herself freely, she would not hold herself accountable for acting on the second-personal address issued to her (SPS 8, 258).

As I declared at the outset, I am less interested in the merits of Darwall's account in itself, than in whether it can help Korsgaard establish morality. Specifically, my interest is in whether we can import Darwall's insights about the way in which the SPS uniquely facilitates the agent's autonomy. In the reminder of this chapter, I explore that question. My conclusion will be that we can't.

6.5.1. The naïve reasoner and autonomy

We have seen that Darwall argues that the agent does not necessarily realise her autonomy in non-second-personal standpoints. Darwall argues for that position through the naïve reasoner. The naïve reasoner is someone who, whilst intelligibly engaged in deliberation, takes it that the sources of her reasons are the objects of her desires. If the naïve reasoner can deliberate under the idea that her reasons are grounded outside her will, then it cannot be the case that she *must* deliberate under the idea of her own normative status to herself - of her own autonomy.

However, the constructivist would reject that conditional. She could accept the characterisation of the naïve reasoner as developed by Darwall. But, she would not think that that characterisation shows that the naïve reasoner does not assume her own autonomy.

The naïve reasoner allows her actions to be determined by something other than the law of her will. But the constructivist would observe that is just how Kant describes heteronomy (GW 4:441).[15] So, for all that Darwall has said, the naïve reasoner could be but an example of heteronomy - heteronomy in the sense of failing to express one's autonomy. So, heteronomy in that sense assumes the agent's autonomy. If the description of the naïve reasoner fits the description of the heteronomous agent, it follows that *as far as that goes* the naïve reasoner might be assuming her own autonomy but failing to express it.

This observation on behalf of the constructivist points out that Darwall's notion of the naïve reasoner does not imply that autonomy is not a necessary presupposition of the standpoint. But it doesn't say that autonomy is being presupposed by the naïve reasoner either. For all that Darwall has said, the naïve reasoner might be assuming her autonomy, or she might not. That is the point of the above observation.

To determine whether the naïve reasoner does or does not assume her own autonomy, the constructivist would dig deeper into the presuppositions adopted by the naïve reasoner. To do that she would attempt to reinforce the premise of the Korsgaard/Kant argument which Darwall challenges - namely, that when deliberating we take the object of our choice to be valuable only because we value it - by digging deeper into the presuppositions of the naïve

[15] A reminder that by 'heteronomous' I mean the failing to express one's autonomy, rather than a description of where the source of one's governance is located. I follow O'Neill's use in her (2003, p.9).

reasoner conceded by Darwall. She could produce a version of Korsgaard's argument for the normative mind, as follows.[16]

We know that the naïve reasoner thinks both that her reasons are grounded outside her will, and that her desires track those reasons. However, once her desires present a presumed reason to her (call it 'R1'), the agent still has to decide whether to act on it or not. For the agent to decide whether to act on R1 or not, she will have to deliberate. For her to count as deliberating at all, she will have to think that the conclusion of her deliberation - her decision - will be delivered by her own deliberation. This much has been granted by Darwall.

But what about the reasons through which the naïve reasoner will determine whether to act on R1? Call those reasons 'R2'. If the naïve reasoner does not assume her own autonomy, she will think that R2 are also grounded outside her will. But in this case, she will also have to decide whether to act on them. For that, she will look to R3, but there too she will have to decide whether to act on R3. We are on our way to an infinite regress.

So, if the naïve agent does not assume her own autonomy – if she thinks that all her reasons are grounded outside her will - she will be committed to an infinite regress in decision making. She could never *decide* what to do. Practical reason seeks to determine what to do. So, the thought that one can never decide, or determine, what to do, cannot be a presupposition of reason. And, if the idea that one can never decide what to do is an implication of thinking that her reasons are grounded outside her will, we must conclude that thinking that her reasons are grounded outside her will cannot be a presupposition of the naïve reasoner's practical reason. So Darwall's picture of the naïve reasoner as someone who is committed to thinking that all her reasons are grounded outside her will, is not, after all, intelligible for the constructivist.

If the idea that the naïve reasoner takes it that R2 (and R3, R4, …, R_n) – i.e. the determining reasons of her decision of whether to act on R1 – are also grounded outside her will leads to an unintelligible notion of the naïve reasoner, then we must think that she takes R2 (or R3, R4, …, R_n) to be determined by her own will. But this is for the naïve reasoner to reason under the idea of her own autonomy. So, if the naïve reasoner accepts the presumed

[16] We saw Korsgaard's argument of the normative mind in Section 3.3.1.

normativity of what she takes to be reasons grounded outside her will, she is being heteronomous.[17]

So Darwall's picture of the naïve reasoner would not convince the constructivist that autonomy is not a presupposition of the practical standpoint. If we are to concede a version of the naïve reasoner that is intelligible to the constructivist, what we get is but a heteronomous agent.[18] As such, the constructivist would not accept that the agent need not assume autonomy from the practical standpoint.

6.6.1. Repositioning the dialectic

Darwall's use of the naïve reasoner has two purposes. First, he uses it to argue that autonomy is not necessarily presupposed from the non-second personal standpoint. This motivates his idea that autonomy is *uniquely* realised within the SPS. Second, having argued that autonomy is not necessarily presupposed from the non-second person standpoint, he disables the Korsgaard/Kant argument for the idea of agents as ends in themselves by pointing out that that argument relies on the assumption that the agent presupposes her own autonomy. In short, Darwall's naïve reasoner is supposed to show that reason can deliver neither the normativity of the moral law, nor its content.

---

[17] Darwall has a supplementary argument to the view that practical reason does not presuppose autonomy of the will. He argues that theoretical reason displays the same freedoms as practical reason, yet we do not think that there is something akin to autonomy of the will in theoretical reason. Darwall's challenge there is why think that practical reason presupposes autonomy whilst theoretical reason doesn't (SPS 214-216). If my discussion in this section of the text is correct - that is, if it is the case that the Korsgaardian constructivist must defend the view that the agent is the source of her own normativity - then the Korsgaardian constructivist would have to meet Darwall's challenge in one of two ways. 1. She could say that the difference between practical reason and theoretical reason is contained in the argument I developed on her behalf. She would have to elaborate that. Perhaps the will is involved differently in practical reason as it is in theoretical reason. Or 2, she could argue that theoretical reason involves something akin to autonomy of the will. Korsgaard seems to hint that she would favour this option in SN 93.

Which option the constructivist would pursue, how she would develop it, and whether it would be successful, are not questions it is within the remit of this thesis to answer. The only point I wish to make is that, if I am right that she has to defend the autonomy of the naïve reasoner through something like the argument I have proposed, she will have to answer Darwall's challenge in one of the ways I have mentioned.

[18] Korsgaard has also noted that, if Darwall's argument for the idea that the freedoms presupposed in the practical standpoint did not amount to autonomy worked, it would also work against his own account. If we can intelligibly think of a non-second person reasoner who thinks that her reasons are sourced outside her will, presumably we can equally think of a second-person reasoner who thinks that the second-person reasons are sourced on something other than the second-person relationship (C. M. Korsgaard 2007). Darwall replies in (S. Darwall 2007).

If the constructivist were successful in reinstating the idea that the naïve reasoner does presuppose autonomy, the Korsgaard/Kant argument for the idea of agents as ends would be restored. However, this would not deliver the moral law. As I have pointed out, the argument I have crafted on behalf of the constructivist as a response to Darwall's naïve reasoner is a version of Korsgaard's argument that the agent necessarily sees herself as an end. We know from the discussion in Section 3.4.3. that, if no further argument is offered, from the argument that one must see oneself as an end we get only a categorical imperative whose scope of universalization is comprised by the agent alone, and not by all agents, as the moral law is meant to be. In Korsgaard's terms, from the presupposition of autonomy by the practical standpoint, we obtain the categorical imperative as the law of the will, but not the categorical imperative as the moral law. That is, we obtain both the normativity and the content of an imperative for the agent to treat herself as an end in herself: a Formula of Humanity for One. So, even if the constructivist's response to Darwall's naïve reasoner worked, we would still be short of a full account of the moral law.

Korsgaard's account of the publicity of reasons is her attempt to expand this Formula of Humanity for One into a Formula of Humanity for All Humanity. But I argued in the previous chapter that she could not forge that link because her account of the way in which we are to treat each other as ends, and thus incorporate others into one's categorical imperative, rules out the expression of the agent's autonomy.

I turned to Darwall's SPS because it promises an account of the way in which we are normative to each other which uniquely facilitates the agent's autonomy. My interest is not with whether the SPS facilitates the agent's autonomy *uniquely*, but with whether it can accommodate it at all. If my arguments on behalf of the constructivist worked, they would deny that the SPS *uniquely* facilitates the agent's autonomy. But they wouldn't deny that the SPS facilitates the agent's autonomy *tout court*. So my interest in Darwall's SPS remains untouched at this point. I now turn to examine whether Darwall's account of the way in which the SPS facilitates the agent's autonomy can be used by the constructivist. I will argue that it can't. The obstacles in the way are similar to the obstacles in the way of showing that the agent can treat another's reasons as reasons without violating the expression of her autonomy.

6.7.1. The SPS and heteronomy

I have explained above that the constructivist would reject the idea that the naïve reasoner does not presuppose her own autonomy. The arguments which I presented on behalf of the constructivist apply to the naïve reasoner in virtue of being a reasoner - an agent. As such, the constructivist would apply similar arguments to all agents. In as much as the second-personal reasoner is an agent, he will also be autonomous. That is to say that the second-personal reasoner will be autonomous prior to engaging in second-personal reasoning. The question then, is whether the SPS allows for the expression of the agent's autonomy. Let me emphasise: the question I am addressing is not whether the agent can express her autonomy in the SPS *as Darwall would have it*. It is rather whether the agent can express her autonomy in the SPS *in the constructivist's terms*, namely, given that the agent is already autonomous outside the SPS.

To recall, the agent expresses her autonomy by ensuring that her maxim is fit as a principle. A maxim is fit as a principle only if the agent can will its universalization without in so doing, placing herself against her own normative status to herself. (It is only the agent's own normative status to herself that matters here because, as I argued in Chapters 2, 3, and 5 that is all that the constructivist argument for autonomy delivers on its own. That is why morality is a challenge for constructivism.) In other words, the agent expresses her autonomy only when her decision for action is determined by her own normativity – when, in her decisions, she treats herself as an end in herself. If the agent's decisions for action are not determined by her own normativity, she will be acting heteronomously. So, the question of whether the SPS allows for the expression of the agent's autonomy, is the question of whether decisions for action taken whist holding the presuppositions constitutive of the SPS can be determined by the agent's own normative status for herself.

One occupies the SPS under two different roles: as the issuer of an address, and as a recipient of an address. Of these, it must be more difficult to defend the expression of the recipient's autonomy, than of the issuer's, as the recipient is the one who is given direction by another. So that is where I will focus my discussion. If it is the case that one cannot express one's autonomy as a recipient of a second-personal address, it is sufficient to show that the SPS is not adequate for the expression of the agent's autonomy.

The SPS has two features which are supposed to ensure that the agent, as recipient of a second-personal address, expresses her autonomy. First, when responding to another's second-personal address, it is within the terms of the SPS that the agent takes herself to be responding to someone else's authority, and not to the objects of her volition. That feature marks the SPS apart from the naïve reasoner. Whilst the naïve reasoner takes it that her reasons are grounded on the objects of her volition, in the SPS the agent necessarily takes it that her reasons (the second-personal reasons to which she is a recipient) have an authoritative source.[19]

According to Darwall, taking your reasons to be grounded on another's authority is different from taking them to be grounded on the objects of your volition. This is because if you take your reasons to be grounded on another's authority, you take yourself to be responsible for acting on the authority of the other - you take yourself to be autonomous (SPS 290). This is not the case if you take your reasons to be grounded on the objects of your volition. If you take your reasons to be grounded on the objects of your volition 'it would simply be impossible freely … to act against [them]' (SPS 290).

I read that to mean that in the first case you are aware that there is a claim on you independent from your interests. That makes you aware that you can *not* act on your interests. This idea harks back to the negative aspect of autonomy as presented by Kant. That is, that the agent sees herself as distinct from her desires (and other mind activities/states).[20] In the second case, by contrast, the agent can't distinguish herself from the activities/states of her mind. As such, she cannot be aware of herself as a source of normativity for herself.

This feature alone does not account for the agent's autonomy. That is because, although the agent takes her reasons to be grounded in another's authority, and not on the objects of her volition, she nevertheless locates the grounding of her reasons *outside her will*. As stated above, the agent expresses her autonomy only if her decisions for action are reached

[19] Of course, I argued that the constructivist would insist that the naïve reasoner does assume her own autonomy deep down. But in so far at some superficial level she thinks that her reasons are grounded outside her will, it serves as a good comparison to the way in which the agent could not think that whilst in the SPS.

[20] We have seen this idea both in the presentation of Kant in Section 5.2.1., and in Korsgaard's own version in Section 3.3.1.

through her own will, or reason.[21] Acting on a claim whose normativity is presumed to be based outside the agent's own will, is just to act heteronomously.

This first feature of the relation between the SPS and autonomy is fortified by the second feature mentioned above. This is that the agent recipient of the second-personal address self-imposes the other's address *freely*, where by 'freely' we mean autonomously. Note, however, that one cannot freely self-impose *any* presumed normative source. For example, the naïve reasoner could not freely self-impose a claim whose presumed normativity is thought to originate in the object of one's desire. She couldn't freely self-impose such a claim because to self-impose a claim is to go against one's own normative status.[22]

The key ideas comprised in the two features supposed to ensure that the SPS is hospitable to the agent's expression of her autonomy are these: that it is not the case that the agent can freely self-impose *any* claim on one's will; that the objects of one's desires and that another's authority are relevantly different sources (or presumed sources) of claims; and that claims originating in another's authority can be freely self-imposed, whilst claims presumed to be originating in the objects of volition can't.

Whether the SPS allows for the expression of autonomy or not, then, turns on the idea that claims originating in another's authority can be freely self-imposed. I argue that this idea does not stand scrutiny.

For me to impose your authority upon myself is for me to allow your authority to govern the outcome of my deliberation. And for me to do so freely is for me to have reached the decision through the application of the law of my will. We are familiar with how this is done. The application of the law of my will consists in ensuring that my maxim is fit as principle. A maxim is fit as a principle only if I can will its universalisation without in so doing setting myself against my own normative status to myself.

For example, if I decide to go to the cinema just on the basis that it is something I want to do, I allow my wants and inclinations to rule me. In so doing, I undermine my own authority over myself. In that sense, I set myself against my own normative status. Or, if I promise you to repay your loan even though I fully expect not to be able to do it, I am

[21] Darwall seems to have this point in mind when he introduces the theoretical benefits of Fichte's Point in SPS p.20.

[22] I expand on this briefly below. For a fuller treatment see Section 5.2.1.

setting myself against myself because, if it is ok for me to take an exception to the rule of making a promise on this occasion, it will also be ok to do so in any future occasion; and if it is ok for me to take an exception to the rule of making a promise on any occasion, then there is no rule. Here I set myself against myself because my maxim is pitched to contradict the very concept I am trying to employ: that of a promise.

So for me to freely self-impose your authority is for me to decide to do so having satisfied myself that the universalization of the maxim in question does not set me against my own normativity. Here we can go in two different ways. We can look at whether I can freely self-impose your authority from a position where I don't see you as a source of normativity for me. Or, we can look at whether I can freely self-impose your authority from a position where I am already subject to it. Let us consider first of all whether I can freely self-impose your authority from a position where I don't see you as a source of normativity for me.

It was precisely the idea that one cannot reason one's way to the idea of others as ends in themselves that motivated Darwall's SPS. That fact gives us a quick answer to the question of whether I can freely self-impose your authority from a position where I don't already see you as a source of normativity for me: that is that I can't, but not because I cannot self-impose it *freely*, but because I cannot self-impose it at all. Darwall and Korsgaard are in agreement that there simply is no rational route to treating you as having a normative status to me unless I see you as having a normative status to me. Where Korsgaard brings in the publicity of reasons to account for interpersonal normativity, Darwall brings in the SPS.

If I cannot self-impose your authority at all – if seeing you as an authoritative source for me is not something I can decide to do – seeing you as an authoritative source for me is not something through which I can express my autonomy or not. Let us next look at whether I can freely self-impose your authority from a position where I am already subject to it.

Darwall is clear on what it is to act on another's second-personal address. It is to do what you say to do just because you ask me to – just because you make a claim on my will (SPS 3, 4). This is just what it is to respond to your reasons as reasons – it is to do what your reasons say to do because your reasons say to do it. We saw this in detail in Section 5.3.3. I argued there that that is similar to deciding to do what my desire says to do just because it places a claim on my will. And that is to allow my decision for action to be determined by something other than my will. Since I couldn't will the universalization of a maxim that posits its normative source in something other than my will without setting myself in

conflict with my own authority, that decision would not express my autonomy. Therefore, in deciding to abide by your authority just because you are authoritative to me – that is, just because you place a direct claim on my will – I would be heteronomous.

We might think that Darwall is shielded from this argument by the key ideas involved in establishing the link between the SPS and the expression of the agent's autonomy which I listed above (on page 141). This is that another's authority and the objects of one's volition are relevantly different source of claims. They are relevantly different with respect to whether the agent can impose those claims upon herself freely.

It is plausible to suppose that my desires and the objects of my volition are *relevantly similar* with respect to whether the agent can freely self-impose claims originating in them. If this is the case, then my argument above would not be valid, for it would draw an illicit analogy between my desires and your authority, as the claims they each place on my will are differently related to the exercise of my autonomy.

However, it is not clear that the idea that the claims originating in your authority and the claims originating in the objects of my volition are differently related to the expression of my autonomy can be defended. Darwall seems to say (on page 140) that awareness of a claim whose normativity is grounded on your authority produces in me the idea of autonomy because it makes me aware that I can act contrary to my interests. In contrast, I wouldn't become aware that I can act contrary to my interests if I thought that my interests always track my reasons. If I'm not aware that I can act contrary to my interests, I can't gain the distance between my mental phenomena and I which leads me to the idea of my autonomy.

In light of the constructivist's argument for her take of the naïve reasoner, that is unsatisfactory on three counts. Firstly, it follows from the constructivist's argument that whenever we face the question of what to do, we assume our own autonomy (Sections 3.4.2. and 5.5.1.). Secondly, according to that sketch of Darwall's view, what matters is what yields the idea of autonomy. And what yields the idea of autonomy are claims which do not conform with my interests and desires.[23] But whether a claim on my will conforms to my interests or not is not determined by the origin of that claim. You can place a claim on my will that conforms to my desires. If I am stepping on your foot because somehow I think

[23] I take interests and desires to be interchangeable, unless otherwise indicated.

that you want me to, and you ask me second-personally to remove my foot from the top of yours, I will gladly do it, as I would like to go somewhere else. Equally, I can have desires which are at odds with each other. I really want to wear that sleeveless dress because it is so beautiful, but I really don't want to be cold. So, we can't draw a difference between claims placed by another's authority and by one's on the basis of whether they concord with my interests or not, and thus yield the idea of autonomy.

Thirdly, it is one thing to have the idea of autonomy, and another to express one's autonomy. Whilst in order to express one's autonomy one has to have the idea, the presupposition of one's autonomy, one might have the idea of one's autonomy and fail to express it – this is what heteronomy consists in. I might get the idea of my own autonomy out of facing a choice between two conflicting interests of mine. But I will fail to express my autonomy if I make my choice on a maxim the universalization of which I cannot will. Equally, it doesn't follow from the fact that a second-personal claim on my will leads me to the idea of my own autonomy that acting on that claim will be an the expression of my autonomy.

Furthermore, if your second-personal claim appears on my will just as my desires appear on my will - namely, without my volitional input - it is even harder to see how your second-personal claims relate to my autonomy and to the expression of my autonomy differently from my desires.

I conclude, then, that we have good reasons to think that my argument that deciding to abide by your authority because it is your authority, is similar to deciding to abide by my desires because they are my desires. Both decisions would be instances of heteronomy.

Trying a different line of thinking – one which we also tried in the previous chapter – we might say that I can freely decide to abide by your authority if I decide that doing what you say passes the universalization test. For example, if you ask me to remove my foot from the top of yours, I might consider doing it because I realise that the unevenness of your foot is the cause of a nagging discomfort and that I would be better off stepping on firm ground. Or I might consider doing it because I judge that it will harmonise relations between us, which is something towards which I endeavour.

If the maxim versions of those considerations would stand the test of universalization, and it is plausible to think that they would, then my decision would be expressive of my

autonomy. As such they would reflect my own normativity to myself. However, the process of reaching that decision leaves your normativity aside. Your practical relevance to me has been to present a possible course of action for me; but that possible course of action does not come imbued with normativity - with your authority. I bestow normativity upon it through *my* decision that the proposed action is the right thing to do. And if I reach that decision properly, i.e. via the application of the test of universalisation, I bestow normativity to it freely.

But it is clear both from Darwall's characterisation of the SPS, and from the role which the SPS is supposed to perform in moral relations that that is not an instance of second-personal reason giving/taking. Darwall's characterisation of the SPS was crafted partly by contrasting it against giving advice (SPS 49), and against acting as an epistemic pointer (SPS 6-7). The contrast was that in both those cases the agent expects her address to suggest to her addressee that it might be a good idea for her to ϕ; whilst in the second-personal address, the agent is telling her addressee to ϕ *because she says so* (SPS 49). But in the example that I have provided, in deciding freely whether your address is what I ought to do, I precisely consider whether it is a good idea for me to do what you say to do. As such I sidestep your authority and instead I decide whether it is what I ought to do through my own reasoning alone.

I conclude that the SPS cannot help the constructivist furnish an account of interpersonal normativity that allows for the expression of the agent's autonomy. Darwall argues that the SPS uniquely facilitates the expression of the agent's autonomy. The constructivist - at least Korsgaard - has to reject the idea that the agent cannot realise her autonomy on her own. That is because one of the main tenets of her constructivism is that the agent is a source of her own normativity. For the SPS to be acceptable to Korsgaard, it has to allow for the agent to realise her autonomy on her own. Once we reinstate the full autonomy of the individual agent, we find that the SPS does not allow for the expression of the agent's autonomy any more than the sharing of reasons advocated by the publicity of reasons does. The reasons why are the same. That is that by definition I express my autonomy by bestowing normativity on the considerations involved in my practical reasoning. But also by definition of the publicity of reasons as well as the SPS, your reasons or address enter my practical reasoning with a normativity of their own. Both these theses are conceptually opposed to one another. That is why abiding by one of those sources of normativity involves overriding the other.

If the problem here has arisen from adapting Darwall's SPS into a constructivist framework, Carla Bagnoli's work presents itself as a promising alternative. That is because she develops an avowedly constructivist account of interpersonal normativity which is supposed to enable the fullest realisation of the agent's autonomy. I present Bagnoli's view, and explore how far it takes us in establishing a constructivist morality, in the next chapter.

- 7 -

Bagnoli's relational autonomy

7.1.1. Situating the discussion

The main purpose of this thesis is to ascertain whether constructivism can accommodate morality. To do so I have taken Korsgaard's constructivist account as my case study. There were two principal reasons for that choice. One is that Korsgaard has crafted one of the most developed constructivist accounts of morality, so engaging with it would be a more fruitful exercise than engaging with less developed accounts. The other reason is that Korsgaard's account is controversial and, in my view, misunderstood in equal measure. By using it as my focus, I can present what I think is a more plausible and more interesting reading.

We saw that Korsgaard's main idea in her attempt to establishing morality is that of the publicity of reasons. Korsgaard's account of the publicity of reasons is supposed to capture the sense in which we are normative to one another. We are normative to one another, she argues, just in virtue of being the self-reflective creatures that we are. The elements of self-reflection are what makes one normative to oneself, and it is also what makes others normative to oneself. In this way, others' reasons, more specifically, others' normative status, becomes a normative status for oneself.

I argued that Korsgaard's account of the way in which others are normative to oneself is at odds with her account of the way in which one is normative to oneself. That is because, I further argued, being normative to oneself requires that one *not* abide by the normative status of others.

I then looked to Darwall's account of the second-person standpoint (SPS). Darwall's account was promising because one of the selling points of the SPS is that it uniquely facilitates the agent's expression of her autonomy. However, I tried to show that the way in which

Darwall argues that the SPS is the only practical standpoint from which the agent can express her autonomy is not acceptable for the constructivist, as it removes the main constructivist tenet that the agent is the source of her normativity. I argued that, once the normativity on the individual agent is reinstated, the SPS cannot allow for the expression of autonomy any more than the sharing of reasons advocated by the publicity of reasons could. In other words, once Darwall's account is made acceptable to the constructivist, the same problems as we encountered in Korsgaard's account reappear, namely, it cannot hold together abiding by one's own normativity and second-personally responding to another's address.

I now turn to Carla Bagnoli's work. In view of the discussion so far, Bagnoli's work provides a promising alternative. Bagnoli's work, like Darwall's, purports to deliver an account of interpersonal normativity which allows for the fullest realisation of the agent's autonomy. But some of the ways in which Bagnoli's account differs from Darwall's makes it especially promising for my project.

Firstly, Bagnoli's argument that autonomy is most fully realised within interpersonal normative relationships involves a reconceptualization of the notion of autonomy. She advances an account of autonomy according to which autonomy is a relational phenomenon. This strategy is interesting because the problem I identified in Korsgaard's account was that of reconciling one's normative status for oneself, i.e. one's autonomy, with the normative status of others. A revised conception of autonomy presents the possibility of a situation where that problem does not arise.

Secondly, whilst the problem with Darwall's account is that making it compatible with constructivism engenders the same difficulties as we find in Korsgaard's account, Bagnoli's account is avowedly constructivist. That means we won't have to impose constructivist requisites upon it, in the way that we did with Darwall's account. If Bagnoli's account of relational autonomy is successful, it will be so within a constructivist framework.

7.1.2 Bagnoli and Korsgaard

As I have just mentioned, the problem I identified in Korsgaard's attempt to establish morality is that, on the one hand, her account of *morality* requires that we abide by the

normative status which according to her account we each have for each other; but on the other hand her account of *normativity* requires that the agent abides by her own normative status to herself. I argued that these two requirements were incompatible with each other. As I explained in 5.2.1., the idea that the agent is normative for herself is captured in the notion of autonomy. To be autonomous is to have a normative status for oneself. So, to abide by one's normative status for oneself is to express one's autonomy. The expression of one's autonomy is exercised through the application of the categorical imperative in one's practical deliberation. Korsgaard calls this the process of 'reflective endorsement'. That is, we endorse a given course of action having reflected on whether it is the right thing to do, and we reflect on whether a course of action is the right thing to do by applying the categorical imperative.[1] Korsgaard's notion of reflective endorsement then captures the way in which the agent expresses her autonomy.

My criticism of Korsgaard's account can be put in the following terms. If we look to reflective endorsement for the standard of correctness of actions, then abiding by the normative status of others counts as wrong action. That is because, I argued, abiding by the normative status of others precludes one's full reflective endorsement.

Bagnoli sees a different problem in Korsgaard's reliance on reflective endorsement to provide the standard of correctness of actions. She argues that it makes Korsgaard's account ill-equipped for diagnosing the wrong of cases of immoralism, such as the Mafioso. According to Bagnoli, this is what prevents Korsgaard's account from producing an adequate moral theory.

In response to the shortcomings she identifies in Korsgaard's account of reflective endorsement, Bagnoli develops an account of autonomy according to which one's autonomy intrinsically involves seeing others as ends. Hers is a relational account of autonomy. Since the problem I raised for Korsgaard's account is that her account of autonomy rules out treating others as end, Bagnoli's account of relational autonomy seems to offer just what would solve that problem. If my autonomy *consists* partly in seeing you as an end, then treating you as an end will be an expression of my autonomy, not a violation of it.

[1] This was discussed at length in Section 3.4.3.

The question I seek to answer in this chapter is whether we should replace Korsgaard's account of autonomy - her account based on reflective endorsement - with Bagnoli's account of relational autonomy. I will argue that although doing so would solve Korsgaard's problem of accommodating the requirement to treat others as ends, the cost would be a diminished normative account.

I begin by presenting in detail the problem with Korsgaard's account, as Bagnoli sees it.

7.2.1. Korsgaard and the Mafioso

The Mafioso is of interest to morality because he is a *principled* criminal. That makes him a problem for rationalist accounts of morality – accounts according to which the requirements of morality are, or are discovered by, requirements of reason. As a matter of course the Mafioso lies, steals, kills, and maims. But he doesn't do those things out of pure greed, nor does he do them impulsively, out of lack of self-control. If he did, he wouldn't pose a problem for rationalist. The rationalist would say that the Mafioso's actions are wrong because greed and impulsiveness and lack of self-control do not stand to rational scrutiny. Instead, the Mafioso does those things because they are requirements placed on him by the social system to which he belongs. That is, the Mafioso adheres to norms. And that suggests that he governs himself. He belongs to a system rigidly structured by rules build around a central code of honour, and he *endorses* his belonging to that system. He is ready to sacrifice his own desires, as well as any moral scruples he might have, for the sake of the demands of his clan.

We can characterise him in Korsgaard's terms of practical identities. In those terms, belonging to that system is a practical identity for the Mafioso, that is, being a Mafioso is a description which is true of him, and it is a way of seeing himself 'under which his life is worth living and his actions worth undertaking' (SN 101). Since he values himself as a Mafioso, he recognises the authority of the rules constitutive of being a Mafioso, and abides by them. In short, the Mafioso appears to conduct his practical life through principles issued to him from a source of authority (Bagnoli 2009, pp.477–479).

The Mafioso poses a challenge for rationalist accounts because these accounts tend to use consistency as their standard of correctness. That is, a decision is correct only if it is

consistent with other decisions, or values, of the agent. But the Mafioso's decisions for action seem to be fully consistent with his values and other decisions. As we have seen in the characterisation above, he does endorse leading his life by the commitments and requirements of his practical identity as a Mafioso (Bagnoli 2009, p.478-479). Examples such as the Mafioso highlight a weakness in these rationalist accounts. That is that consistency requirements concern primarily which values or commitments we may have in relation with other values or commitments, but not which values or commitments we must have *tout court*.

Korsgaard's account is a rationalist one, but it has an advantage over rationalist accounts which rely merely on consistency. As we saw in Chapter 3, she argues that there is a value which we must have and that is fundamental to all others. That value acts as the fixed point of reference for consistency requirements. It is one's human identity. With that fixed value in place, it is no longer the case that consistency sets can go in just any direction. They are anchored in one's human identity - in one's valuing of one's own humanity. Ultimately, they have to be consistent with the requirements of the human identity.

I explained that this brings the further challenge of showing how one might value others as ends (Section 2.2.1.). Korsgaard's answer is to argue that valuing oneself 'somehow involves' valuing others (SN 132). So the human identity doubles up as the moral identity and thus it provides the standards of correctness for how to treat others as well as oneself in one's choice of actions. In this way whether an action, or commitment, is right will be a matter of whether it is consistent ultimately with one's moral identity.

With that tool in hand, Korsgaard's brand of rationalism is better equipped than others to diagnose the wrong of the Mafioso. The Mafioso's actions are wrong because his practical identity as a Mafioso is not consistent with his moral identity. Although the reflective endorsement behind his Mafia-related actions expresses his practical identity as a Mafioso, it goes against his more fundamental practical identity as an agent. By endorsing a practical identity which contradicts his fundamental moral identity, he positions himself against his own normative standing to himself, and thus his agency is diminished (Bagnoli 2009, p.479).

So, the purported reflective endorsement of his identity as a Mafioso and all that involves, is not really reflective endorsement, but a corrupt version of it. That is because reflective endorsement, according to Korsgaard, is the way in which the agent expresses her autonomy. Since her autonomy is comprised in her moral identity, to go against her moral

identity is to fail to express her autonomy. Since the Mafioso's actions and commitments go against his moral identity, he fails to express his autonomy. So, his endorsement of his actions and commitments *qua* Mafioso are not fully reflective endorsements. Since agency is the process by which the agent becomes the author of her actions, and, on this view, to become the author of one's actions is to fully reflectively endorse them, then the Mafioso's agency is diminished by less than fully endorsing his actions and commitments.

Bagnoli agrees that the Mafioso doesn't enjoy full agency, but she argues that Korsgaard's diagnosis does not quite capture the way in which it is wrong. That is because according to Bagnoli successful agency is not a matter of ensuring that one's commitments reflect the agent's valuing of her own moral identity. That is, it is not a matter of expressing one's autonomy, which is comprised in one's moral identity. Rather, according to Bagnoli, successful agency is a matter of the agent having the correct representation of herself.

We'll see what an agent's correct representation of herself comes down to in the next section. Before that, let's look at why Bagnoli thinks that the Mafioso's diminished agency is not a matter of his contradicting his moral identity (Bagnoli 2009, pp.480, 485).

Bagnoli points out that a criterion of success for the rationalist diagnosis of the wrong of the Mafioso is that the wrong is shown to be internal to the Mafioso (Bagnoli 2009, p.485). This is because rationalism implies that ethical and moral requirements can be discovered through reasoning alone. If the requirements of morality can be discovered by reason alone, then the rationalist must be able to show to the Mafioso through reason what is wrong with his being a Mafioso. Showing that the wrong of the Mafioso is internal to him would entice the Mafioso to change his ways on pain of irrationality.

Korsgaard diagnosis would meet this condition because it can point out to the Mafioso that his practical identity as a Mafioso contradicts his moral identity. Since his moral identity is his most treasured practical identity, we would show him that in being a Mafioso he is going against his most treasured values, including his own value for his own normative status to himself.[2] If Korsgaard's account is correct, if we presented it to the Mafioso, he would gain a clear picture of the ways in which he is wrong.

[2] He would go against his own normative status for himself because the moral identity is supposed to comprise the normative status of all agents, including his own.

However, Bagnoli argues that the Mafioso does not have a moral identity. So, although Bagnoli agrees that his commitments as a Mafioso go against the value of humanity comprised in *our* human identity, those commitments do not go against the value of humanity comprised in *his* human identity because, Bagnoli argues, he doesn't have one. As such, there is no internal cost for the Mafioso, and so no rational way of convincing him to abandon his practical identity as a Mafioso. In other words, he doesn't have a reason to change his Mafioso ways. So Korsgaard's account fails to meet the internal condition.

I shall explain why Bagnoli thinks that the Mafioso doesn't have a moral identity in a moment. Before that it is worth pausing for a moment to outline the dialectical route which Bagnoli takes from this point.

Korsgaard differentiates between valuing oneself, as comprised in one's autonomy and in one's human identity, and valuing others. She attempts to forge a link between the two which would yield her moral identity. The moral identity would comprise both one's valuing of oneself and of others. We can put her argument schematically like this:

1. Valuing oneself involves valuing others.
2. Everyone must value themselves.
3. Therefore, everyone must value others.

If we think that the Mafioso does not have a moral identity, we might use that to reject premise 1, and thus to show that Korsgaard's attempt to link valuing oneself with valuing others fails. The possibility of the Mafioso, on this view, would be in consonance with the discussion in Chapter 4 of the ways in which Korsgaard's link between valuing oneself and valuing others comes loose. But this route would leave the Mafioso with his own self-value as comprised in his human identity.

However, Bagnoli takes a different route. Instead of using the claim that the Mafioso does not have a moral identity to reject premise 1, she uses it to reject premise 2 - the idea that everyone must value themselves. In this way, she retains the thought that valuing oneself involves valuing others, that is, premise 1. That premise, as we shall see, is a corollary of her account of relational autonomy, according to which the moral identity and the human identity can't conceptually come apart in the way in which they do in Korsgaard's account. (I explain Bagnoli's account of relational autonomy below.) If the human identity and the moral identity can't come apart, then, if the Mafioso does not have a moral identity, he

doesn't have a human identity. So, an implication of Bagnoli's dialectical turn is that she challenges Korsgaard's claim that the human identity is necessary and fundamental to all valuing. I will discuss this implication later; for now, my purpose is to signal that that is the direction which the discussion will be taking.

With that dialectical signpost in place, I know look at why Bagnoli thinks that the Mafioso does not have a moral identity.

7.3.1. Morality and value

In a nutshell, Bagnoli argues that the Mafioso does not have a moral identity because having that identity involves valuing oneself and others in a certain way, and the Mafioso values neither himself nor others in that way. In this section I present Bagnoli's account of the kind of valuing involved in the moral identity, and in the next section I explain why Bagnoli thinks that the Mafioso does not value himself or others in that way.

Although Bagnoli does not overtly express adherence to Korsgaard's account of practical identities, in particular, of human and of moral identities, it is implicit in her discussion that she agrees with them, or at least that she doesn't disagree. So I shall proceed to outline her argument using the notions of the human and moral identities.

In order to get a picture of what kind of valuing is involved in the moral identity, let us recall what having that identity consist in. According to Korsgaard, the moral identity derives from the human identity. Having a human identity consists in seeing oneself as a source of value, *and* to value oneself under that conception. To *see* yourself as a source of value is to see yourself as normative to yourself, or as an end in yourself.[3] So, to *value* yourself conceived of as a source of value is to value yourself as a normative source for yourself, or as having the status of an end in yourself for yourself. So, having a human identity consists in valuing yourself as an end in yourself.

If the human identity consists in conceiving of oneself as a normative source to oneself and of valuing oneself under that conception, the moral identity consists in conceiving of oneself as *one amongst other* normative sources for oneself and of valuing oneself under that

[3] This construal of 'seeing' someone as an end harks back to the one I outlined in Section 5.3.1.

conception.[4] The moral identity then obtains when I see others as sources of normativity for me - as sources of value, or ends in themselves - just as I see myself. When I conceive of others as normative for me, and I conceive of myself as normative for me, those conceptions become unified as a self-conception in which I see myself as normatively on a par with others. This is the moral identity. So, both the human identity and the moral identity consist of *ways* of valuing oneself, and, in the case of the moral identity, of ways of valuing others: as ends in themselves.

For Korsgaard, then, valuing oneself as an end in oneself and valuing others as ends in themselves are two different concepts picking two different phenomena. As we know, she tries to bring the two together through her notion of the publicity of reasons. Through that notion she argues that what makes one normative for oneself, makes others normative for oneself, and oneself normative for others. But the two ideas remain distinct. And the same goes for the human and the moral identities. The two identities can converge into the moral identity, but they remain distinct. Specifically, the human identity remains the most fundamental identity.

Bagnoli, by contrast, develops her account of the type of valuing involved in the moral identity by appeal to Kant's distinction between self-esteem and self-respect. She characterises self-respect and self-esteem in the following way:

> [S]elf-respect results from an appropriate relation to others as having equal standing, it does not depend on the judgment of others. The normative source of self-respect as well as respect of others is mutual attribution of authority. Self-respect is tantamount to the experience of the limitations that the recognition of others put on us. The normative source of self-esteem, instead, is placed on others as bestowing a particular judgement of merit on us. Esteem is of some particular qualities and merits, while respect is the appropriate consideration of one's value independently of any quality or merit (Bagnoli 2009, p.484).

I believe that we can read that characterisation as saying that relating to others/oneself in terms of respect is tantamount to seeing others/oneself as ends in themselves; and that relating to others/oneself in terms of esteem is tantamount to seeing others/oneself in terms of their instrumental status. But to see that takes some unpacking. So let's do that, beginning with respect.

[4] Korsgaard puts it this way in SN 132.

Bagnoli says that respect is a relation which does not depend on the judgement of others. To get an idea of what a relation that does not depend on the judgement of others is like, it helps to consider what a relation that *does* depend on the judgment of others is like. A relation which depends on the judgement of others would be one in which the parties involved evaluate and assess each other, and conclude that they each and each other are worthy of that relation. Friendships, for example, have an element of judgement in their foundation. Whilst a friendship develops in part by the involuntary favourable response by each of the parties involved to the others, before committing to the friendship the parties will also assess the others for attributes important to them. I might find you charming and witty, but if you are a racist or a sexist, I might decide that I don't want to pursue a friendship with you. Being interviewed for a job, or for entry to an exclusive club would, on this view, be paradigmatic examples of relationships overtly based on having a certain attributes.

If this is what it is like for a relation to depend on the judgement of the parties, then for a relation to *not* depend on the judgement of the parties will be for that relation to obtain regardless of what the parties think of each other. It is puzzling to think that I can have a relation with you without my making *any* judgement about you - we might think that I must at least judge that you are *a person*, and not an automaton or a mirage. But Bagnoli gets to this point when she says that relations of respect are formed on the basis of 'mutual attribution of authority' (Bagnoli 2009, p.484, quoted above). I take it that to attribute authority to another we must see them as persons. If my relation of respect to you is based on our mutual attribution of authority, and our mutual attribution of authority contrasts with judgements about each other, it would seem that the attribution of authority is not dependent on any judgement other than the recognition of each other as persons.

Bagnoli also says that 'Self-respect is tantamount to the experience of the limitations that the recognition of others put on us' (Bagnoli 2009, p.484, quoted above). If for something S to put a limitation on me is for S to set an end for me; and for S to put a limitation on me is for me to see S as an end in itself; then the idea that 'the recognition of others' puts limitations on us is tantamount to the idea that to recognise others is to see them as ends in themselves. So, a relation of respect is a relation in which the parties see each other as ends in themselves just in virtue of seeing each other as persons.

If that is what respect is, self-respect will be the same but addressed to the self. That is, it will be a relation that the agent has to herself as an end in herself - as a source of normativity to herself. Finally, self-respect arises from the recognition that others are ends in themselves for oneself - that others are normative for oneself.

Relations based on esteem, in contrast, are relations based on judging others as described above in the examples of friendship, job interviews, and exclusive club membership. Regarding another with esteem is not a matter of recognising her status as an end in herself, rather it is a matter of valuing her just in as much as she exhibits certain contingent and variable features. To esteem someone, then, is to value them instrumentally. And so, to esteem oneself is to value oneself instrumentally. One is not valuable in oneself, one is valuable in as much as one displays some valuable attribute. Self-esteem, like self-respect, arises from the agent relating to others in terms of esteem.

In brief, Bagnoli's use of Kant's distinction between self-respect and self-esteem tells us that relations of respect are relations based on the recognition of each other as ends in themselves; that relations of esteem are relations based on representing each other as instrumentally valuable; and that those ways of relating to others yield a corresponding way of relating to oneself, namely, they yield self-respect and self-esteem respectively.

This exposition also enables us to appreciate why Bagnoli thinks that the human identity and the moral identity cannot come apart. The human identity consists in the agent valuing herself as an end in herself, and the moral identity consists in the agent valuing herself *and* others as ends in themselves. If relating to someone in terms respect is a matter of recognising their status as ends in themselves, then there is a superficial way in which the notion of self-respect maps onto the notion of the human identity. But according to Bagnoli's discussion, the agent's self-respect results from the agent relating to others in terms of respect. So, if self-respect is tantamount to one's human identity, and self-respect necessarily involves respecting others, then one's human identity necessarily involves respecting others – seeing others as ends in themselves. But this is what the moral identity consists in: valuing oneself and others as ends in themselves. So, the human identity collapses into the moral identity. In Bagnoli's account, the two identities cannot come apart any more than self-respect and respect for others can.

7.4.1. The Mafioso and morality

If the human identity and the moral identity are merged in this way, it follows that if the Mafioso does not have a moral identity, he doesn't have a human identity either. Bagnoli goes on to argue that the Mafioso does not have a moral identity by arguing that he doesn't relate to either himself or others in terms of respect. And she argues that the Mafioso does not relate to others or himself in terms of respect by showing that he relates to others in terms of esteem. To do this, Bagnoli draws on sociological studies of the Italian Mafia to extract a picture of the values which sustain and help constitute the structure of the Mafia.[5]

The network of values of Mafia groups revolve around a code of honour. This code of honour is rigidly maintained and enforced through a system of sanctions and incentives. Loyalty to the code is rewarded, and disloyalty is severely punished. The bindingness of the code of honour is grounded on this system of incentives and sanctions. Correspondingly, authority is attributed to different members of the group to the measure of their readiness – or their reputed readiness – to impart those rewards and sanctions, particularly the latter.

Mobsters might nevertheless develop special bonds of kin or friendship. However, the Mafioso sees everyone as bound by that system of incentive and rewards, so he expects everyone to deceive when fluctuations in their circumstances make it propitious to do so. This makes his special personal relations saturated with suspicion, secrecy, and distrust. In short, Mafia members value others according to their reputation for the firm implementation of the code of honour. And, seeing themselves as part of that network, they value themselves also to the extent that they believe themselves to have accrued that reputation.

We might think that this analysis of the value system of the Mafia shows that the features which were thought to pose a challenge to the rationalist, had in fact been misconstrued. One of those features is that the organisation of the Mafia seemed to be neatly governed by alliance to the code of honour, but Bagnoli's analysis reveals that their internal social features around the implementation of the Code make it impossible for the mobsters to ever be safe. That is, being a Mafioso entails not only doing bad things to others. Doing bad things to others by following fully endorsed norms was how we initially presented the challenge which the Mafioso poses to rationalist accounts. Now it turns out that being a

[5] The following is a synthesis of Bagnoli's description over pp. 482-485 of her 2009.

Mafioso makes it more likely than not that one's life will be dominated by fear, before it ends brutally and prematurely.

We might think that this extra information about the Mafioso puts it within reach of rationalist accounts. For now the rationalist could say that the wrong of the Mafioso is that his commitments are self-destructive. However, this take by the rationalist remains unsatisfactory because of the fact that the Mafioso *reflectively endorses* that way of life, costs included. As Bagnoli suggests, for all that reflectively endorsing that way of life for oneself raises questions about the political and social environment in which these people find themselves, it remains a challenge to the rationalist, as everything within the Mafioso's practical landscape seems to be in order. So, there remains reason to believe that the notion of reflective endorsement remains ineffective to diagnose the wrong of the Mafioso. Let us then return to Bagnoli's development of an alternative diagnostic tool.

Having argued that having a moral identity involves relating to oneself and others in terms of respect, and presented a sociological sketch of the organisation of the Mafia, Bagnoli next argues that that sociological account shows that the Mafioso does not relate to oneself and to others in terms of respect. That sociological account shows that the Mafioso assigns value to himself and to others in relation to their reputation as strict guards of the code of honour. To value oneself and others on the basis of reputation is an instance of relating to oneself and to others in terms of esteem. To esteem someone is to value them on the basis of a judgement as to whether they possess certain attributes. The Mafioso values himself and others just to the extent that they are reputed to dispense rewards and, particularly, sanctions as part of the group's code of honour. So, the Mafioso values himself and others in terms of esteem.

If the Mafioso relates to himself and to others in terms of esteem, he doesn't value humanity as such. Therefore, he doesn't have a moral identity. And since Bagnoli merges the human identity with the moral identity, it follows that he doesn't have a human identity either. This is why, according to Bagnoli, Korsgaard's diagnosis of the wrong of the Mafioso does not hit its target. Korsgaard's diagnosis, recall, is that the Mafioso's commitments contradict the value of humanity as comprised in his moral identity. That is supposed to be a wrong because in Korsgaard's account to contradict the value of humanity is to contradict the agent's dearest and most fundamental valuing. However, if the Mafioso does not value

humanity, he is not going against his dearest and most fundamental valuing. Therefore, his wrong cannot be located in his commitments.

7.5.1. The Mafioso and autonomy

According to Bagnoli, the wrong of the Mafioso is not that he goes against his moral identity. Rather, she argues, his wrong is that he lacks a moral identity (Bagnoli 2009, pp.480, 486). And, a corollary of Bagnoli's argument that the Mafioso lacks a moral identity is that he lacks autonomy.

To be autonomous is to take oneself as being normative to oneself. To take someone as being normative to oneself is to relate to them in terms of respect. So, to take oneself as being normative to oneself is to relate to oneself in terms of self-respect. But, according to Bagnoli, to relate to oneself in terms of self-respect is to relate to oneself *and* others in terms of respect. So, according to this account, to see oneself as normative to oneself is to see oneself *and* others as normative to oneself. So, to be autonomous is to see oneself *and* others as normative to oneself. So, autonomy is a relational concept. But to see oneself and others as normative to oneself is comprised in the moral identity. Since, according to Bagnoli, the Mafioso lacks a moral identity, it follows that, on her account, he lacks autonomy.

This, according to Bagnoli, is why the agency of the Mafioso is diminished. He doesn't see himself as an authoritative source, and that is ultimately self-defeating (Bagnoli 2009, p.486). Instead of seeing himself as a *source* of authority, he sees his authority at the mercy of fortune.

Bagnoli takes the weakness of Korsgaard's account to be revealed by the fact that it cannot show that the wrong of the Mafioso is internal. As I explained earlier (Section 7.2.1.), this is a condition which the rationalist accounts must meet. That is because if the requirements of morality are discovered by reason, as would be the case if rationalism were true, they must be accessible to all rational beings. If Bagnoli's account is to be superior to Korsgaard's, her account will have to be able to show to the Mafioso that he has a reason to abandon his Mafioso identity. This might seem a tall order. If the Mafioso lacks autonomy, he is not a full blown agent. As such, it is unclear how his practical landscape might be impinged upon by pointing out to him that his actions and commitments prevent him from becoming a full-

blown agent. Bagnoli admits that news of her diagnosis in itself will not give the Mafioso such a reason. However, her account points at the way in which *we* – moral agents – can alter his practical landscape in such a way that the Mafioso has a reason to change (Bagnoli 2009, pp.477, 488 ff).

Since according to Bagnoli's account, autonomy consists of seeing oneself as one amongst other sources of normativity for oneself, and one comes to see oneself as a source of normativity as a result of seeing others as sources of normativity for herself, we can encourage the development of the Mafioso autonomy by relating to him as a source of normativity, that is, in terms of respect. If we relate to him with respect we are thereby expressing our own autonomous agency. To express our own autonomous agency is, in this account, to treat ourselves and others as sources of normativity for us. So, in treating the Mafioso with respect, we are enabling him to recognise our normative status. Since, according to Bagnoli's account, one recognises one's own normative status by recognising the normative status of others, we would be enabling the Mafioso to develop his own autonomy. And once he begins to develop his autonomy, he will have a reason to continue en route to enhance his agency (Bagnoli 2009, pp.477, 488 ff).

This, then, is Bagnoli's alternative account to Korsgaard's of why the Mafioso's agency is diminished: it is because he lacks autonomy; not because he doesn't ensure that his commitments are consistent with his moral identity. Furthermore, and importantly for us, her account of why the Mafioso's agency is diminished is based on a different understanding of what it takes to enjoy full agency. Whilst for Korsgaard enjoying full agency is a matter of ensuring that one's decisions are in compliance with the value of humanity, i.e. reflective endorsement; for Bagnoli it is a matter of *representing* oneself and others as sources of authority (Bagnoli 2009, pp.478, 487).

If Bagnoli's account is successful we have a redefined conception of autonomy. That is of interest for us because the problem I identified in Korsgaard's attempt to establish morality was that we couldn't reconcile the requirements of autonomy with the requirements of morality. That is because autonomy as construed by Korsgaard requires that one treats oneself as an end in oneself, and morality requires that one treats others as well as oneself as ends in themselves. I argued that rationally treating oneself as an end in oneself rules out rationally treating others as ends in themselves, and vice versa. This problem seems to be avoided by Bagnoli's account. Since according to Bagnoli's account, one sees oneself as an

end in oneself when one encounters the normative limitations placed by others, then treating others as ends in themselves is necessary for autonomy. So, in contrast with Korsgaard's account, on Bagnoli's account, abiding by the normative status of others, far from contradicting one's own autonomy, reaffirms it. If Bagnoli's account works, it looks like we will have found what we were looking for: a way of establishing a constructivist morality.

7.6.1. Bagnoli vs. Korsgaard

My purpose in this chapter is to see whether we could replace Korsgaard's account of autonomy with Bagnoli's and thus avoid the problem I raised to Korsgaard's account - namely, the problem of accommodating the requirement to treat others as ends. In order to decide whether to do so we should get clear on what is at stake between them. Bagnoli and Korsgaard agree that the wrong of the Mafioso is that he lacks full agency. But they disagree on what it is about the Mafioso in virtue of which he lacks agency. According to Korsgaard, the Mafioso lacks agency because his commitments go against his moral identity. Since his moral identity include his autonomy (single autonomy, in Korsgaard's account), by endorsing commitments which go against his moral identity he endorses commitments which go against his autonomy. By endorsing commitments which go against his autonomy, he fails to treat himself as an end in himself. So, according to Korsgaard, the Mafioso lacks full agency because he fails to *treat* himself as an end in himself.

According to Bagnoli, in contrast, the Mafioso lacks full agency because he fails to *see* himself and others as ends in themselves. The notion of the agent as an end in herself is of course the notion of autonomy, and Bagnoli and Korsgaard also disagree on what that entails. According to Korsgaard one's autonomy consists essentially in one seeing oneself as an end in oneself, as a source of normativity for oneself. For one to see oneself *and* others as ends in themselves, reasons have to be public. But this is an add-on to her account of autonomy. Her account of autonomy as the agent seeing only herself as a source of normativity takes conceptual and phenomenological priority. (The failings both in her account of the publicity of reasons, and in the extent to which the publicity of reasons can show that one would conceive of oneself in that way need not concern us at this stage of

the discussion.[6]) According to Bagnoli, one's autonomy essentially involves seeing others as ends in themselves.

So, the debate is on two fronts: one is about what constitutes autonomy; the other is about what constitutes agency. Korsgaard holds that autonomy is constituted by seeing only oneself as an end in oneself, and that agency is constituted by treating oneself as an end in oneself - by expressing one's autonomy. Bagnoli on her part holds that autonomy is constituted by seeing oneself *and* others as ends in themselves, and that agency is constituted by having autonomy.

In order to decide whether we should accept Bagnoli's account and thus avoid the problems at establishing morality which face Korsgaard, we should see how each of Bagnoli's and Korsgaard's account fare on these debates. I start by looking at the debate about what constitutes agency, where agency is construed as the process by which one becomes the author of one's actions.

According to Korsgaard, what is constitutive of agency is the expression of one's autonomy. That is, to become the author of one's actions is to *treat* oneself as an end in oneself. According to Bagnoli, what is constitutive of agency is having autonomy.[7] That is, to become the author of one's actions is to *see* oneself as an end.[8] Exploring the relative merits of both accounts will involve delving into their different accounts of autonomy. However, the first part of the discussion does not turn on which conception of autonomy is at play. For the sake of clarity, then, I will leave the fact that Bagnoli and Korsgaard each have different conceptions of what is constitutive of autonomy out of the initial part of my enquiry. The matter of the different conceptions of autonomy becomes germane to my discussion later on. I will indicate it when it is so.

---

[6] The problems regarding the extent to which the publicity of reasons can show that one would come to see others as ends in themselves were discussed in Chapter 4. The problems concerning the ability of the account of the publicity to establish morality were introduced at length in Chapter 5, and have, of course, been driving the later chapters of the thesis.

[7] Since the expression of one's autonomy implies one's autonomy, we might think that according to Korsgaard, both being autonomous *and* expressing one's autonomy are jointly constitutive of agency. If this were the case, and if Bagnoli were right that the Mafioso lacks autonomy, Korsgaard would be able to diagnose his wrong as lacking autonomy just in the same way as Bagnoli does. However, for Korsgaard autonomy is a transcendental condition for agency, rather than a constituent of it. So, if it is the case that the Mafioso lacks autonomy, Korsgaard would be unable to diagnose his wrong, as Bagnoli argues, because he wouldn't count as an agent.

[8] Once again, the difference in the text between 'seeing' and 'treating' someone as an end, is aligned to my characterisation in Section 5.3.1.

In order to get clearer about what benefits Korsgaard's account offers, as well as in the interest of the discussion to come, it will help to recall why she thinks that the expression of one's autonomy is what is constitutive of agency. With that aim in mind, I provide a sketch of the main claims of her view. It will be only a sketch of the main claims, rather than the arguments supporting those claims, both because the arguments for the view were presented at length in Chapter 3, and because the main claims alone are all we need at present to identify the benefits of the view. I begin by sketching what the expression of autonomy consists of, and then I explain why that is supposed to be constitutive of agency.

The expression of autonomy consists in treating oneself as normative to oneself. One is normative to oneself in virtue of the two commitments expressed in the posing of the question of what to do: the commitment to finding out what the right thing to do is; and the commitment to doing what the right thing to do is. So, one treats oneself as an end in oneself - as a normative source to oneself - by taking those commitments to fruition.[9] Those commitments are taken to fruition by finding out what the right thing to do is, and by doing it because it is the right thing to do. The right thing to do is what the agent decides she can choose to do without losing her identity, ultimately, without losing her identity as an agent. So, the expression of one's autonomy consists in reaffirming one's own identity as an agent.

That expression of one's autonomy is supposed to be constitutive of agency in that the expression of one's autonomy is effected in the choice of actions. In electing one's actions as the expression of one's autonomy, one ensures that one's own source of normativity (i.e. oneself conceived of as an agent) continues through one's actions.[10] This is what it is for one to be the author of one's actions, and thus what constitutes agency.

One advantage of Korsgaard's view is that it is especially well equipped to diagnose the wrong of heteronomous actions.[11] Heteronomous actions are those through which the agent

[9] There is a sense in which Korsgaard's account of autonomy might also be called 'relational'. After all, autonomy is the way in which the agent is normative to herself, and Korsgaard endorses a relational account of normativity. However, it is not relational in the sense that it obtains only when other agents are involved, as is the case with Bagnoli's, as for Korsgaard autonomy is an intrinsic feature of a person.

[10] This is another way of putting the idea that one preserves one's practical identities through action (Section 3.4.1.).

[11] A reminder that I follow O'Neill's construal of heteronomy as contrasting with the expression of autonomy, and not as contrasting with autonomy (O'Neill 2003, p.9). If instead I employed a construal of heteronomy which contrasts with autonomy, then I would rephrase my talk of heteronomous actions in the text as actions through which the agent fails to express her autonomy. I believe this would not affect the main thrust of my discussion.

fails to express her autonomy - she fails to treat herself as an end in herself. Those actions are wrong because in choosing them the agent fails to ensure that that her source of normativity continues to be normative through her actions. That is the way in which her agency is unstable.

The ability to diagnose the wrong of heteronomous actions is something which it would appear Bagnoli's account lacks. If agency is constituted by autonomy, as Bagnoli's account has it, all it takes for one to be an agent is that one be autonomous. If we retain the ambition of diagnosing wrong actions as actions which destabilise one's agency, then so long as one is autonomous one's agency is in order, and one's actions are correct, or at least permissible. But heteronomous actions, or actions which *fail* to express the agent's autonomy, assume the agent's autonomy. So, actions which fail to express the agent's autonomy could not be classified as wrong, according to Bagnoli's account, because one would still be autonomous.

On the other hand, if it is the case that the Mafioso does not have autonomy, then Bagnoli's account can diagnose his wrong, whilst Korsgaard's can't. Korsgaard's account can't diagnose the wrong of lacking autonomy because, according to Korsgaard, wrongness consists in failing to express one's autonomy. If one has no autonomy, one cannot *fail* to express it. That is, one who lacks autonomy cannot express autonomy. But that won't be a failing. So Korsgaard's account cannot diagnose the *wrong* of lacking autonomy, whilst Bagnoli's can.

But it looks like Bagnoli's account can't do much more than that. Suppose that, in accordance with Bagnoli's account, we treat the Mafioso with respect and he develops autonomy. Suppose further that he continues on his old ways: he continues to steal, kill, and main, only now he is autonomous. Since according to Bagnoli's account, having autonomy is constitutive of agency, and this Mafioso is now autonomous, Bagnoli's account does not seem to be able to diagnose what his wrong is.

We might be tempted to think that Bagnoli could make room for Korsgaard's view at this point. Perhaps she could hold that both having autonomy and expressing one's autonomy are jointly constitutive of agency. She would have to find some justification for taking that view, so as for it not to be ad hoc, but it is plausible to think that such justification might be forthcoming from the requirements of agency. If Bagnoli could do that, her account would be considerably empowered. It would be able to diagnose the wrong both of the non-

autonomous and of the autonomous Mafioso. This would put her account above Korsgaard's, since the latter can only diagnose the wrong of the autonomous Mafioso.

However, this option is not open to Bagnoli. To see why, we need to bring into the discussion Korsgaard's and Bagnoli's different conceptions of autonomy. The reason why Bagnoli cannot embrace Korsgaard's view that the expression of autonomy is constitutive of agency is that Korsgaard's view implies a non-relational account of autonomy. That is, Korsgaard's view of the constitution of agency is tied up to her account of the constitution of autonomy. According to Korsgaard's view one expresses one's autonomy by bringing to fruition the commitments manifested in the posing of the question of what to do. Those commitments do not arise from relating to anyone, but just from one's own reflective nature (Section 3.3.1.). Since those commitments are constitutive of autonomy, it follows that one's autonomy arises from one's own reflective nature. There is no *need* for the agent to see herself as one amongst other ends in themselves. This is why I said earlier that Korsgaard's attempt to bring others into one's normative landscape is an add-on to her account of autonomy.

As far as the current discussion goes, then, it looks like the stakes are as follows (the stakes as far as the whole thesis goes are outlined below): If we accept Korsgaard's account, we are able to diagnose the wrong of heteronomous actions, but we are unable to diagnose the Mafioso if he does lack autonomy. If we accept Bagnoli's account, we are able to diagnose the Mafioso if he does lack autonomy, but we are unable to diagnose the wrong of heteronomous actions.

However, it follows from Korsgaard's view that the Mafioso *does* have autonomy. As we saw, according to that view, autonomy is constituted by a relation the agent stands to herself. The parts of the agent which stand on this relation come into relief in the posing of the question of what to do. One poses the question of what to do whenever one becomes aware of an inclination, or desire in oneself. According to Korsgaard, that is just what it is to be a self-reflective being. Unless we were prepared to deny that the Mafioso lacks self-reflection at that elementary level - that he doesn't wonder whether to go on a diet or to continue enjoying rich desserts; how to spend his evening off; or even how to dispose of his latest corpse - if we accept Korsgaard's account, then we accept that the Mafioso is autonomous. And if he is autonomous, we can diagnose the wrong of his heteronomous actions as actions which defy his own normative status to himself - his own autonomy.

Bagnoli argues that the Mafioso lacks autonomy on the basis that he lacks self-respect. But if self-respect is a matter of treating oneself as a normative source, as earlier I explained it is, then if we accept Korsgaard's account, lacking self-respect will be construed as a type of heteronomy. So, if we accept Korsgaard's account, by showing that the Mafioso lacks self-respect, we just show that he displays heteronomy.

If this discussion is correct, then Korsgaard's account gives us more than Bagnoli's: Korsgaard's account can diagnose the wrong of heteronomous actions, whilst Bagnoli's can't; and although Korsgaard's can't diagnose the lack of autonomy as a wrong, and Bagnoli's can, it follows from Korsgaard's account that all persons are autonomous in as much they are self-reflective. So although Korsgaard's account cannot diagnose the wrong of lack of autonomy, it can diagnose all actions of persons.

I looked at Bagnoli's account to see if her relational account of autonomy could help Korsgaard's account avoid the problem I have identified in her account - that is the problem of reconciling the ideas of being subject to one's one normative status, and of being subject to the normative status of others. If autonomy were relational in the way advanced by Bagnoli, that problem would be avoided because according to that account to be normative to oneself is but an aspect of being subject to the normativity of others.

However, I have now argued that a relational account of autonomy is at odds with the way of diagnosing the wrong of heteronomy which follows from Korsgaard's account. That is because Korsgaard's way of diagnosing the wrong of heteronomy implies that autonomy is primarily non-relational.

As for how this discussion fits within the whole thesis, the stakes are as follows: If we adopt Bagnoli's account of relational autonomy, we are able to account for morality, but since we lose the ability to diagnose the wrong of heteronomous actions (at least if we want to be able to retain agency as providing the standard of correctness), we are rendered unable to diagnose the wrong of all those wrong actions performed by moral agents, rather than by immoral ones, such as the Mafioso would be in Bagnoli's account.

If, on the other hand, we keep Korsgaard's account, we are able to account for the wrong of heteronomous actions, but if my arguments in Chapter 5 are correct, then we are unable to build a moral account. In that case, all the heteronomous actions we can account for are those who go against the agent's own normative status to herself, and *not* those who go

against the normative status of others. Moreover, I have argued that actions which reflect the normative status that others have for the agent are deemed heteronomous in Korsgaard's account. So if we accept Korsgaard's account over Bagnoli's, Korsgaard's account remains unable to diagnose the wrong of the Mafioso. But that is not because the Mafioso lacks autonomy, as Bagnoli argued – according to Korsgaard account, he does have autonomy – but because his autonomy requires that he dismisses the normative claim which others make on him.

I conclude that constructivism as presented here lacks the ability to combine a satisfactory account of normativity with the requirement to treat others as ends.

7.7.1. Overall conclusion

My task in this thesis has been to assess whether metaethical constructivism can meet the two main challenges it faces. These challenges are 1) to show that its account of normativity can deliver thoroughgoing standards of correctness without collapsing into a form of either realism or expressivism, and without appealing to implausible premises; and 2) to show that it can accommodate a morality based on the idea of treating persons as ends.

In Chapter 1 I argued that constructivism can meet the first challenge. However, in the rest of the thesis I have argued that, at least it its Korsgaardian version, its account of normativity is conceptually opposed to the requirement to treat others as ends. In this last chapter, I have argued that Bagnoli's attempt to unite the requirements of normativity with the requirement to treat others as ends result in a deficient account of normativity.

I conclude that an account of morality based on treating persons as ends cannot be accommodated within the conceptual framework of constructivism because it cannot combine a successful account of normativity with the requirement to treat others as ends.

Works cited

Allison, H.E., 1986. Morality and Freedom: Kant's Reciprocity Thesis. *The Philosophical Review*, XCV(3), pp.393–425.

Bagnoli, C., 2011. Constructivism in Metaethics. In E. N. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/win2011/entries/constructivism-metaethics/ [Accessed March 16, 2012].

Bagnoli, C., 2002. Moral Constructivism: A Phenomenological Argument. *Topoi*, pp.125–138.

Bagnoli, C., 2009. The Mafioso Case: Autonomy and Self-respect. *Ethical Theory and Moral Practice*, 12(5), pp.477–493.

Darwall, S., 2007. Reply to Korsgaard, Wallace, and Watson. *Ethics*, 118(1), pp.52–69.

Darwall, S.L., 1985. *Impartial Reason* New ed., Cornell University Press.

Darwall, S.L., 2006. *The second-person standpoint : morality, respect, and accountability*, Cambridge, Mass. ; London: Harvard University Press.

Darwall, S.L., 2009. Why Kant Needs the Second-Person Standpoint. In *The Blackwell Guide to Kant's Ethics*. pp. 138–158.

Enoch, D., 2006. Agency, shmagency: Why normativity won't come from what is constitutive of action ,. *Philosophical Review* ,, 115 ,(2 ,), pp.169–198 ,.

Enoch, D., 2011. Shmagency revisited. In *New Waves in Metaethics*. pp. 208–233.

Gert, J., 2002. Korsgaard's private-reasons argument. *Philosophy and Phenomenological Research*, LXIV, pp.303–324.

Gibbard, A., 1999. Morality as consistency in living: Korsgaard's Kantian lectures. *Ethics*, 110, pp.140–164.

Herman, B., 1983. Integrity and Impartiality. In *The Practice of Moral Judgment*. Harvard University Press, pp. 23–45.

Herman, B., 1992. What happens to the consequences? In *The Practice of Moral Judgment*. pp. 94–112.

Hill, T.E., 1983. Self-Respect Reconsidered. In *Autonomy and Self-Respect*.

Hill, T.E., 1973. Servility and Self-Respect. In *Autonomy and Self-Respect*.

Hume, D., 1975. *Enquiries concerning Human Understanding and concerning the Principles of Morals* 3rd ed. L. A. Selby-Bigge, ed., OUP Oxford.

Kant, I., 1998. *Kant: Groundwork of the Metaphysics of Morals*, Cambridge University Press.

Korsgaard, C. M., 2007. Autonomy and the second person within: a commentary on Stephen Darwall's The Second-Person Standpoint. *Ethics*, 118, pp.8–23.

Korsgaard, Christine M., 2009. *Self-constitution : agency, identity, and integrity*, Oxford ; New York: Oxford University Press.

Korsgaard, Christine M., 2008. *The constitution of agency : essays on practical reason and moral psychology*, Oxford ; New York: Oxford University Press.

Korsgaard, Christine M., 1993. The reasons we can share: an attack on the distinction between agent-relative and agent-neutral values. In *Creating the Kingdom of Ends*. pp. 275–310.

Korsgaard, Christine M., 1996. *The Sources of Normativity*, Cambridge ; New York: Cambridge University Press.

Lenman, J., 2010. Humean Constructivism in Moral Theory. In *Oxford Studies in Metaethics vol 5*. pp. 175–194.

Nagel, T., 1986. *The view from nowhere*, New York: Oxford University Press. Available at: http://www.loc.gov/catdir/enhancements/fy0638/85031002-d.html.

O'Neill, O., 2003. Autonomy: The Emperor's New Clothes. *Proceedings of the Aristotelian Society*, 77, pp.1–21.

Ridge, M., 2005a. Agent-neutral vs. agent-relative reasons. *Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/reasons-agent/.

Ridge, M., 2005b. Why must we treat humanity with respect? Evaluating the regress argument. *European Journal of Analytic Philosophy*, 1, pp.57–74.

van Roojen, M., 2011. Moral Cognitivism vs. Non-Cognitivism. In E. N. Zalta, ed. *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/spr2011/entries/moral-cognitivism/ [Accessed March 24, 2012].

Smith, M., 1994. *The moral problem*, Oxford: Blackwell.

Strawson, P.F., 1962. Freedom and resentment. In *Freedom and Resentment and Other Essays*. pp. 1–25.

Street, S., 2008. Constructivism about Reasons. In Oxford Studies in Metaethics. pp. 207– 245.

Street, S., 2010. What is constructivism in ethics and metaethics? ,. *Philosophy Compass* ,, 5 ,(5 ,), pp.363–384 ,.

Velleman, J.D., 2009. *How We Get Along* 1st ed., Cambridge University Press.

Wallace, R.J., 2009. The publicity of reasons. *Philosophical Perspectives*, 23(1), pp.471–497.

Williams, B., 1980. Internal and External Reasons. In *Moral Luck*. pp. 101–113.

Williams, B.A.O., 1981. *Moral luck : philosophical papers 1973-1980*, Cambridge: Cambridge University Press.

Acknowledgements

I am enormously grateful to my supervisors, Michael Ridge and Elinor Mason, for their tireless dedication; for their support, patience, and encouragement; and for their tact and sensitivity over the last stretch of writing. At times I couldn't believe how lucky I was to have them. I am also grateful to Jonathan Hearn, who supervised the very early stages of this thesis too. His kindness and encouragement at that point were crucial.

I would also like to thank my dear friend Hannah Dawson. All those afternoons spent in conversation did not help much towards getting the thing written, but they did so much to sharpen the ideas and nourish the soul. Thanks also to Nicole Hall-Elfick for her boundless kindness and support, and for the many, many conversations about beauty and other important matters. Thanks too to Clare MacCumhaill for her friendship, support, and good humour. She never, ever fails to inspire me. I would like to thank Morwenna Griffiths for the wonderful coffee breaks, and for giving me the right advice at just the right time and just in the right way. This thesis would be very different if it weren't for her. Thanks also to Eric Freund for being the best illustration of a good will I know, and to Vlad Eatwell for attending to dinner whilst the drama came to his kitchen once more.

I would like to thank my parents, Maribel and Jesús, for their love, patience, and unwavering support, and for the euros. My beloved sister Beatriz has performed the best job of being a sister anyone could wish for during this time. I can't wait to repay it.

Finally, thanks to Matthew Nudds for much, much more than I can even begin to say.