The history and viability of constitutivism in ethics

Adam J Roberts

Thesis for the BPhil Philosophy

Abstract

The topic of this thesis is constitutivism in ethics: its history and its viability. Constitutivists hope to show that moral standards are normative because if we are not following them, we are failing to be agents. We cannot escape trying to be agents: it is just our way of trying to do anything in particular. In failing to be agents, then, we are failing to do what we ourselves are trying to.

The force of normativity, on a constitutivist view, turns out to be the force of our own efforts to do things. Moral or pro-moral standards are what we must be guided by for our own attempts to do things to be internally consistent. The immoral would-be agent, then, is one who is divided against themself. They are trying to do the impossible: to be agents in a way that agents cannot.

Defending the viability of constitutivism – one half of my topic – only takes defending one form of it. This thesis focuses on Kantian constitutivism, which may be the most promising kind of it. It focuses in particular on Christine Korsgaard's version of that position, as presented in *The Sources of Normativity*, *Self-Constitution*, her recent *Fellow Creatures*, and a range of essays.

The four substantive chapters of this thesis set out to answer four different questions. The first is what the components of Kantian constitutivism are. Two of those four components are claimed to be characteristic of constitutivism in ethics, and the chapter points to the historical background of all four of them. The second question is in what sense we face the problem of agency. The chapter first distinguishes rational and nonrational agency. It then explains how, on the view, rationality is self-conceptual, and attempts to pin down this talk of problem-facing.

The final two chapters of this thesis take up two closely related questions. These are why our moral psychology should feature what Kant calls "incentives," and why, as Kant and Aristotle believe, it should feature not just acts, but ends. It argues that on the Kantian constitutivist picture, there are two demands on a moral philosophy of action. Those demands can be seen to generate the functional concepts of incentives and, via the steps of a Kantian moral argument, ends.

Contents

Acknowledgements

Sometimes a piece of work stands on its own, as its author's first foray into a field. Other times, that work is just the tip - if a 30,000 word tip - of an ongoing series of efforts. This thesis falls into the latter category. I began trying to make sense of Christine Korsgaard's position while studying at King's College London, taking a module on human dignity with Andrea Sangiovanni. Many of the objections we were hearing to Korsgaard's position seemed to make it out to be so obviously misguided that she would surely have noticed, the Harvard professor that she is.

I had a similar experience on a module with Thomas Pink, about the nature of normativity. For both of those modules at King's, I wrote essays in partial defence of Korsgaard's arguments. Not because of the teaching, for which I am grateful, neither essay was particularly well-received. The same was the case when I moved to Oxford and began to write on Korsgaard for the BPhil. Tom Sinclair alerted me to how much I still had to fully understand about her position. Had I heeded his warning, I might not have heard it again from the BPhil examiners. Adapting an argument of Korsgaard's into political philosophy, my very first assessed essay here earned me a fail mark.

Help understanding Korsgaard's metaethics and aspects of her philosophy of action came in the following terms from Ralf Bader, my second supervisor. As John Callanan had earlier in a module at King's which included Kant's humanity argument, Ralf also helped me to better understand Korsgaard's debt or relationship to Kant. From Ralf's, Tom's, and John's supervision, I am grateful to have benefited. In my final two terms here, for this thesis specifically, I had the supervision of Alison Hills.

I am grateful never to have left a supervision with Alison without renewed self-confidence and a clearer vision. The many mistakes and imperfections of this thesis remain either because I have stubbornly kept them in, or because I failed to prepare drafts and redrafts in time to clutter her inbox with them. For supporting me through the process of writing up this thesis, I am grateful to my families, natural and chosen. Of those chosen, I would single out Mati Regent, Ivy Manning, and Adam Lewis-Douglas. They could not quite keep me sane, but they kept me closer than I otherwise would have been. I hope I either have been or will be able to return the favour.

This thesis' reviewer 2 remarked that "the candidate's efforts to defend the viability of constitutivism show at least that s/he has thoroughly mastered (one is even tempted to say 'internalized') the consistency and coherence of Korsgaard's version" of it. If that three-quarter-compliment and this thesis' overall distinction mark are fair, it is only because of people like those mentioned here. The drawing of the tree of Ténéré on page ii is Mati's, and used with his permission.

1 Introduction

1.1 A brief overview

1.1.1

The topic of this thesis is constitutivism in ethics: its history and its viability. Constitutivists hope to show that moral standards are normative because if we are not following them, we are failing to be agents. We cannot escape trying to be agents: it is just our way of trying to do anything in particular. In failing to be agents, then, we are failing to do what we ourselves are trying to.

The force of normativity, on a constitutivist view, turns out to be the force of our own efforts to do things. Moral standards are what we must be guided by for our own attempts to do things to be internally consistent.[1] The immoral would-be agent, then, is one who is divided against themself. They are trying to do the impossible: to be agents in a way that agents cannot.

Defending the viability of constitutivism – one half of my topic – only takes defending one form of it. I intend to focus on Kantian constitutivism, what I believe is the most promising kind of it. I will be focussing in particular on Christine Korsgaard's version of that position, as presented in *The Sources of Normativity*, *Self-Constitution*, her recent *Fellow Creatures*, and a range of essays.

[1] I use "moral standards" in a loose sense here, to include those standards which encourage or press us towards morality (2.1.4).

1.1.2

Let me begin by summarising what it means to be a Kantian constitutivist. This is not meant to be a satisfying summary: think of it more as an elevator pitch. The idea is that it might slightly demystify that part of my topic as I begin to talk about its significance and my approach to it.

Kantian constitutivists argue that we conceive of our attempts to act – to make things happen – in such a way that proposing to act immorally turns out to be a way of proposing something logically impossible. In a strict way of speaking, an immoral action is a contradiction in terms. Something cannot both be an action and be anything but morally permissible.

As such, acting immorally is not strictly speaking something we can choose to do. Therefore, it is a mistake to see it as an option when considering what to do. None of this prevents us from talking in a loose way about our acting immorally, and of course, we often seem to. When trying to choose our actions, however – from that practical perspective – it is a mistake to see immoral options as open.

1.1.3

The structure of this thesis will be as follows. I will begin by explaining the significance of my topic. As I see it, Kantian constitutivism – a position which is both metaethical and normative – is more promising than realism or noncognitivism or any first-order theory built upon them (1.2.1–1.2.2). I will explain why I nevertheless describe my optimism about Kantian constitutivism as cautious (1.2.3), and then what I see this thesis as contributing to Korsgaard's work (1.2.4). Before closing out this introduction, I will explain my approach to the two aspects of my topic (1.3).

The four substantive chapters of this thesis will set out to answer four different questions. The first is what the components of Kantian constitutivism are. Two of these I believe are characteristic of constitutivism in ethics (2.1). The latter two make what I see as the most viable constitutivism Kantian (2.2). I will explain how the four components are related, and how they fit together into one complete view (2.3). I will then move on to metaethical questions, and the relation between constitutivism and constructivism (2.4). Along the way in this chapter, I will be relating the ideas I cover to ones in history, pointing to Plato and the Stoics (2.1.2–2.1.3), Hobbes and Descartes (2.2.3), and of course Kant himself.

In the second of these four substantive chapters – chapter three – I will take up the task of making clearer than much of the literature does in what sense we face the problem of agency. I will begin by distinguishing between rational and nonrational agency (3.1.2, 3.1.7), explain how rationality is self-conceptual (3.1.6), and attempt to pin down the ubiquitous talk of problem-facing (3.1.3–3.1.5). I will attempt to break down the senses in which the problem of agency is inescapable (3.2), before returning to Plato and Kant to demystify the claim immoral actions are not actions strictly speaking (3.3).

The final two chapters of this thesis will take up two closely related questions. These are why our moral psychology should feature what Kant calls "incentives," and why, as Kant and Aristotle believe, it should feature not just acts, but ends. I will begin to answer the first question by returning to the appeal or the promise of constructivism (4.1). I will argue that on the Kantian constitutivist picture, there are two demands on a philosophy of action (4.2). I will then argue that there is a problem seeing how we can determine our actions' content, and that part of the solution must be that the content is proposed to us as Kant and Korsgaard believe (4.3).

The final chapter, as I said, turns to ends. I will sketch Korsgaard's version of a moral argument which is supposed to show that the only intelligible way to act is by Kant's categorical imperative (5.1.1–5.1.2). That argument ends, specifically, with Kant's Formula of Universal Law. I will make a different argument, leading to Kant's idea of a kingdom of ends. I will argue that we cannot intelligibly consider only the content of our own incentives (5.1.3–5.1.4), but that our incentives being reconcilable requires them being backed by chains or branches of ends (5.2). Those branches can be brought together into a tree, or Kant's kingdom, of ends (5.3.1). I will close by considering what action (5.3.2) and immorality (5.3.3) end up being.

1.2 Significance of the topic

1.2.1

As I said above, I will begin to grapple with my topic by explaining its significance: why you ought to care about it, at least as I see it. Of course, one reason to be interested in any philosophical theory is if you think it might work. I am hopeful – a different attitude to confident – that Kantian constitutivism, of Korsgaard's kind, works. That is, I am cautiously optimistic that it accurately describes a problem we are faced with, and then how to solve it.[2] Cautious optimism might not seem a striking or compelling attitude to argue for in a thesis. It is, however, a more positive attitude than I believe is warranted for any alternative.

[2] This way of describing what it means for a theory in philosophy to work might sound relativistic or antirealist. It is neither of those things: rather, it is neutral in those respects, and appropriately. It is a substantive question whether all of us face the same, or similarly structured, problems. As such, it is a substantive question whether a theory which is helpful to you, or works for you, could be helpful to everyone. It is equally a substantive issue whether the aim of all the problems which we face as human beings is to describe some problem-independent realm of truths about the nature of the world or what to do.

Kantian constitutivism, as I will go on to explain, is a form of Kantian constructivism (2.4.1). As I understand it, it is the form which Kantian constructivism takes in ethics, both normative and meta. There could be other Kantian constructivisms in other fields of philosophy: John Rawls used that label for his political philosophy.[3] In epistemology or metaphysics, Henry Allison's understanding of Kant's idealism might be called Kantian constructivist.[4]

In ethics, Kantian constructivism or constitutivism is sometimes contrasted with realism and noncognitivism. Its proponents argue that it rises above the unavoidable inadequacies of those other kinds of theories. I believe there are good arguments that both realist and noncognitivist kinds of theory are inadequate necessarily. By comparison, then, being a cautious optimist about Kantian constitutivism does mean something.

By realism, I mean the view that there are moral truths, and ones which "really" hold, not just from some perspective. The problem with realism, as I see it, is that there cannot be an interesting "real" kind of normativity. Suppose that you keep asking a realist, as you always can, why it is you ought to do what they say you should. Ultimately, their only answer can be that you *just do*: it is some kind of hard fact that you should. Since the realist's explanation ends there – they have run out of "why" answers – it can only be left puzzling what this kind of fact is meant to be, metaphysically, or what it means to say that it is normative.

The Kantian constitutivist has a better answer. What it generally means for something to be normative for you is for it to be the only way of doing something which you are trying to do. There are standards which are normative categorically if there are things you cannot help but

[3] See e.g. Rawls, 'Kantian Constructivism in Moral Theory.' For what Rawls' and Korsgaard's views have in common, as both Kantian constructivist, see COA 317–24.
[4] This is obviously not a claim I can defend here. I made it previously in my 'The missing two-thirds of metaethics.' The idea is that what Allison calls "epistemic conditions" are the theoretical equivalent of what, in this thesis, I will call "practical conditions" (2.2.4). For Allison's account of the former, see e.g. his *Kant's Transcendental Idealism*, 11–19.

try to do, where those standards just describe the way of doing what you cannot help but try to do. There is nothing puzzling about what normativity is, on this picture. It is the force of your own willing to do things. The reason why it presses us, as if it were outside of us, is that we can fail to grasp how to do what we are trying to. It is not because normativity is something external, something real, directed at us.

In this way, Kantian constitutivism is not a realist view. There are not normative facts which hold "really," outside of the perspectives we take up in trying to do things (2.4.3–2.4.5). Inside of those perspectives, however, normative facts are as real as anything. The normativity of what we are trying to do is granted to us within our trying to do it. Kantian constitutivists take this perspective talk seriously, and it is crucial to how they see the distinctiveness of their metaethics. Different metaethical answers, they believe, are warranted inside of and outside of certain perspectives. It would be difficult to explain this at this moment, without first having more fully explained Kantian constitutivism. It is a topic I will come back to, however, at a number of points in this thesis.

1.2.2

The problem with noncognitivism is, in a way, the same as the problem with realism. Like the realist, the noncognitivist fails to explain what it means for something to be normative. For the noncognitivist, however, that failure is willing. Perceiving the inadequacy of a realist metaethics, the noncognitivist gives up on normativity as a fact about, or as a property of, ways of being or acting.

Pinning down exactly what it means to be a noncognitivist is a little complicated, as easy as the sanitised, textbook concept is to state. Noncognitivists deny semantic factualism: on their understanding of normative claims, such claims – like, "you ought to do this" – are not

assertions of facts, or that something is the case. Instead such claims express norms you accept, or emotional responses to the "this," or public plans you subscribe to, or similarly nonfactual, noncognitive things. Nevertheless, we find Allan Gibbard asking:

> Does this mean that there are no facts of what I ought to do, no truths and falsehoods? Previously I thought so, but other philosophers challenged me to say what this denial could mean. In this book, I withdraw the denial and turn non-commital. In one sense there clearly are "facts" of what a person ought to do, and in a sense of the word 'true' there is a truth of the matter. That's a minimalist sense, in which "It's *true* that pain is to be avoided" just amounts to saying that pain is to be avoided – and likewise for "It's a fact that". Perhaps, as I used to think, there are senses too in which we can sensibly debate whether *ought* conclusions are true or false. Nothing in this book, though, depends on whether there is any such sense.[5]

Being a noncognitivist, then, we might say is not about the surface semantics of normative claims, but about their ultimate analysis: whether they resolve into assertions of fact, or into attitude expressions. The problem with noncognitivism, I said, is that it gives up on normativity as a fact or a property. We can now be more specific: it does that giving up at that level of claims' ultimate analysis.

The reason why that giving up on normativity as a fact is problematic is that we, as practical reasoners, are not able to give up on it in that way. We all find ourselves faced, in some sense inescapably, with the problem of working out what to do (3.1–3.2). In order to make progress

[5] Gibbard, *Thinking How to Live*, x.

with that problem, we need principles or procedures which can guide us: we need something to narrow down all the options we have, or show us the way forward with them.

Realism and Kantian constitutivism both attempt to do that narrowing down for us. When they claim that certain principles are the right ones, they mean for us to then be guided by them. Realist theories recommend certain principles as the ones we *just ought* to try to follow. Kantian constitutivism attempts to show us that following certain principles is how to do what we are already attempting.[6] Noncognitivist theories, on the other hand, officially give up on that project of helping us see what to do. That is, on pain of failing to be noncognitivist, they cannot tell us that any action solves our problem. They can *reveal* to us that we have certain norms or plans, but cannot tell us to act on them. If they did, they would be realists or constructivists about the value of the norms to which we already subscribe.

Of course, the noncognitivist may respond that this is the task of first-order moral theory. Running together normative ethics and metaethics is at most something Kantian constitutivists want to do. This response, however, misunderstands the problem. The problem is that a noncognitivist metaethics *makes* first-order theories unable to strictly tell us to act in any given way.

This may not seem like the most obvious problem with noncognitivism, but I think it is the one that matters. As Sharon Street puts it, both the constructivist and the noncognitivist "charges the other with nothing less than having overlooked or lost sight of the central task of metaethics."[7] A way of seeing what noncognitivism does is give up on the problem which realism wants to work for. For realists, as for Kantian constructivists, the problem is the

[6] For the difference between realism and constructivism framed this way, see COA 315–26 and SC 64–67.
[7] Street, 'What is Constructivism in Ethics and Metaethics,' 375.

practical one: what to do. Noncognitivists instead design theories to work for the problem of understanding our moral practice: what we do. This is a different, theoretical problem, and answering it does not answer the practical one. Knowing you are fated to only express your values does not help you choose what values to have.

These are not meant to be careful, uncontroversial challenges to realism and noncognitivism. I am, more than anything, trying to explain my own view of the significance of my topic. I have been suggesting that if, as I am, you are comparably pessimistic about the other options, then being cautiously optimistic about a theory in ethics can mean something.

1.2.3

I have not explained what it is about Kantian constitutivism that makes my optimism about it cautious. The simple answer is that I am not confident in my understanding of the issues here. Kantian constitutivism is systematic, and the problems of moral philosophy can be as hard to pin down as to answer. There are more overlaps, loose ends, and ambiguities than I will have thought of, let alone take up here.

Near the beginning of Book X of the *Republic*, Plato has Socrates ask of Glaucon:

> 'Can you tell me what the production of imitations is, in general? I ask because I'm not entirely clear in my own mind what it thinks it is.'
>
> 'As if I'm going to be clear about it,' he said, 'if you aren't!'
>
> 'There'd be nothing strange in that,' I said. 'People who don't see so well often see things before the sharper-sighted.'

> ‘That’s so,’ he said; ‘but if anything does occur to me, with you here I certainly won’t be rushing to say so. You do the looking!’ (Rep 595c–6a)

There is merit here in Glaucon’s not “rushing to say so.” I do have higher hopes about Kantian constitutivism than about any alternative. At the same time, many serious philosophers have doubted, and do doubt, the ideas which in this thesis I will be trying to make plausible. They have doubted them both separately and in combinations, as constitutivism and the Kantian kind of it.

1.2.4

That was not the only reason to bring in Glaucon. That quote which I just gave ends with him saying: “You do the looking!” What the historical thread of this thesis aims to bring out, of course, is that others have already done the looking. Moreover, the four components I will be describing come together in a contemporary Kantian constitutivism like Korsgaard’s, which raises the question what gap my thesis fills: why more than Korsgaard’s own books and essays is needed.

In describing how “philosophers at present do not go in much for origin stories” in ethics and politics, Korsgaard comments that:

> Analytic philosophy these days has become a crisp no-nonsense discipline, aligning itself with the sciences rather than with literature, and rejecting any modes of understanding whose methodological credentials are obscure. Since philosophy is a discipline of self-understanding, we are of course right to try to understand our own methods where we can. But crisp no-nonsense attitudes often express nothing more than a lack of imagination, and a desire to eliminate perplexity as soon as possible. (REM 2)

One hallmark of analytic philosophy is the making of conceptual distinctions which distil ideas into smaller ones and bring out their premises. Part of what this thesis aims to do is to go through that process with Kantian constitutivism: breaking it down into its parts, noting their history, and piecing them together again. It will do the same with some concepts along the way: for example, as I said, I will be arguing that we can distinguish between at least three senses in which the problem of agency is inescapable (3.2.2).

To think that what Korsgaard argues can be clarified is not to say her own work is particularly unclear. I find her style rewarding to read, and, cautiously, I do find her arguments promising. As Thomas Nagel, a realist, puts it, "her project is of the first importance" whether or not her own answers to the questions which she grapples with succeed (SN 209). I will not just be clarifying Korsgaard's work, however. As I said, in chapters four and five of this thesis, I will be attempting to motivate two aspects of Korsgaard's philosophy of action. Those aspects – incentives and ends, respectively – need that motivating, both in the sense that their validity needs defending and in the sense that these parts of Korsgaard's argument are missing.

What I mean by this is that Korsgaard's first-order moral arguments, particularly in the *Sources* and *Self-Constitution*, seem to depend or be based on a particular moral psychology. Korsgaard can make it sound as if that moral psychology is the starting point of her argument. However, I do not believe that needs or ought to be the case. The task of chapters four and five – a third of this thesis – is to give that psychology an explicit grounding.

1.3 Approach to history and viability

1.3.1

I have been explaining the significance of the topic of this thesis, or why one might care about it. I now want to be more explicit still about how I will approach the history and the viability of constitutivism. As I have been suggesting, I will approach its viability primarily by detailing a Kantian kind of constitutivism in a way which makes it plausible that it is coherent and well-grounded, and in that sense viable.

In his 'Shmagency Revisited,' following up the critique of constitutivism he presented in his 'Agency, Shmagency,' David Enoch complains that "Korsgaard and her followers, it often seems, are just not that much into responding to objections."[8] In the case of small or superficial misunderstandings of Korsgaard's views, the evidence is not, I think, on Enoch's side here. A range of essays and their footnotes acknowledge the ways in which Korsgaard has been pushed to defend her position.

Where Korsgaard may not seem to be so forthcoming is where objections to her views are based on somewhat deeper misunderstandings. There is a reason of sorts for this, which is that Korsgaard's views are deeply systematic. When your views form a system in the way that Korsgaard's or, famously, Kant's own views do, the only way to respond to the deepest objections is to restate your view. You can try to do that more clearly, in a way that makes the parts which are most relevant spring out, so to speak; but there is no shortcut to responding to such objections, or no difference between responding to them and stating your view. Korsgaard describes herself as "tempted to think of a philosophical system as a kind of

[8] Enoch, 'Shmagency Revisited,' 218.

complex and intricately structured machine, like an engine" (RK 1165). Nothing short of an explanation of the whole of an engine may persuade the unconvinced that it will not explode.

Enoch's own "shmagency" objection may be an example of one based on deep and even systematic misunderstandings. Since plenty has been written on it elsewhere, I will only have one brief claim to make about it here.[9] Enoch's deepest misunderstanding, I think, relates to some issues I have already covered. It is that he takes the constitutivist to be trying to offer the same kind of normativity as the realist. A moment ago, I suggested that realistic kinds of normativity are uninteresting (1.2.1). In asking for a reason, *tout court*, to be agents, however, Enoch must be asking for that kind of normativity. Constitutivists are only trying to show morality is normative within our trying to be agents. That normativity matters, however, since we are trying to be agents in trying to do anything (3.2).

As I have been saying, the contributions I will be seeking to make on constitutivism's viability are an analysis of its components, a clarification of the problem it addresses, and a two-part argument for Korsgaard's action theory. As I said at the beginning, all it takes to defend the viability of constitutivism in ethics is to defend one kind of it. These contributions, I believe, address the most important missing parts of a defence of Kantian constitutivism. The approach which I will not be taking, at least directly, is the one which Enoch suggests is needed. I will not be focusing on mopping up the objections in the literature to Kantian constitutivism.[10] As I have just been suggesting, however, I do not see a sharp distinction between responding to objections and restating a theory. My emphasis will be partly on clarifying the theory, and partly on filling in the missing parts of it.

[9] I have previously taken up Enoch's "shmagency" objection in my 'Why be a Kantian agent,' and in my 'In defence of Korsgaard's arguments for human and animal value.'
[10] The two works I mention in the previous footnote were attempts by me to do this.

1.3.2

Viability is only half of my topic. The approach I will be taking in this thesis to history will be topical, not chronological, bringing in the background to ideas in the process of explaining them. The history of the ideas which come together in Kantian constitutivism is a particularly rich one. Part of that history is obviously Kantian, although there is *always* room to disagree how much.

Kant's own ideas did not spring out of nowhere, so to speak spontaneously. His moral philosophy in particular owes a great deal to the Stoics. We see Kant acknowledging his relationship to this history at points, such as when in the *Religion* he praises how the Stoics "derived their universal moral principle from … freedom," as he also tries to (Rel 6:57).

In *Self-Constitution* and an earlier essay, Korsgaard traces some of the connections between Kant, read as a constitutivist, and Plato's *Republic* (SC 133–184; COA 100–26). She argues that both Kant and Plato saw moral standards as ones by which we have to be guided in order to be, strictly speaking, acting at all or agents. Korsgaard also connects both Kant's theory of normativity and his philosophy of action to Aristotle, arguing that both saw normative standards as internal to things, and both thought we choose ends when we choose actions (COA 174–206, 215–229; SC 8–18, 27–44).

There is plenty of scope both to challenge and develop Korsgaard's readings of constitutivism's history. For example, I think it less clear than Korsgaard seems to that Plato in the *Republic* shares the narrow concern of contemporary constitutivists with agency. As Luca Ferrero puts it, the contemporary view is that "agency is a very special enterprise," being "distinctively 'inescapable,'" and this is what "sets [it] apart from all other enterprises and

explains why constitutivism is focused on it."[11] At points, however Plato seems as concerned with passive robustness as with agency. We see this when has Socrates talk, for example, about how an oligarchic person changes "when one of two warring elements in him receives outside support" (Rep 559e).

Moreover, there is scope to say a great deal more about what Kantian constitutivism shares with, if not owes to, the likes of the Stoics and the early constructivists. By the early constructivists, I mean the likes of Descartes and Hobbes, who have been read as having a kind of problem-solving philosophy (2.2.3). In the case of the Stoics, some amount has been written on such topics as the comparability of their and Kant's headline moral standards and accounts of dignity. There continues to be plenty of space, however, for comparisons of the Stoics' views to Kant's read explicitly as Korsgaard reads them, as those of a constitutivist.

[11] Ferrero, 'Constitutivism and the Inescapability of Agency,' 308.

2 What are the components of Kantian constitutivism?

2.1 How morality applies to and binds us

2.1.1

I have been explaining the significance of the topic of this thesis and my approach to its subject's history and viability. The task of this chapter is to broach the four components which I am claiming come together in Kantian constitutivism. Kantian constitutivists, I said earlier, argue that we conceive of our attempts to act – to make things happen – in such a way that proposing to act immorally turns out to be a way of proposing something logically impossible. As such, it is a mistake to see acting immorally as an option for us. As an option, if not as something which can happen, the idea of an immoral action is incoherent.

I believe that this position can be helpfully broken down into four components, the first two of which are shared with any constitutivism in ethics. Were those components to need one-word names, they might be: "Agency," "Possibility," "Practicality," and finally "Indefeasibility."

2.1.2

The first of those components, common to any constitutivism in ethics, is that agency is the locus or the site of moral standards. This, as I mentioned, is an idea Korsgaard finds in the *Republic*, and is also one which I believe that Kant found in the Stoics. For Plato's Socrates, "those who are totally corrupted and perfectly unjust are also perfectly incapable of action," unable to do anything as whole people or city-states (Rep 352c–d). For the Stoics, the moral

action is the one we would perform if our souls were not moved by outside forces, "in accordance with our own human nature as well as that of the universe" (LEP vii 88).

Agency for Plato, in other words, is what we have when our parts work together as something more, as a whole. For the Stoics, it is what is left when nothing is allowed to take the reins and do our acting for us. We find versions of both pictures in Kant. The autonomy of the moral agent consists in their being more than, and bringing order to, their inclinations. At the same time, that choice to bring order to their inclinations and what Kant calls their "incentives" must be thought of as not based on things, but rather as made spontaneously (5.1.1).

On all the pictures at which I have just gestured, agency and morality are supposed to be closely related. The way that moral standards relate to us is somehow through the nature or the problem of agency. As I said earlier, this focus on agency is common to all versions of constitutivism, with the idea being that agency has some "very special" features, such as being "distinctively 'inescapable'."[12] Later on in this thesis, I will explain what that kind of inescapability is, and why it matters (3.2). Roughly, the idea is we are trying to be agents whatever particular things we are trying to do. Trying to be agents is not some further, perhaps optional project in addition to our trying to cook a meal or be successful. It is our *way* of trying to do anything, and so all of our tryings to do have it in common.

2.1.3

The second of the four components I mentioned, and again one which comes together in any constitutivism, is that the way in which moral standards bind us is by conditioning our possibilities. That is, the way that immoral actions are ruled out for us on the view is by being ruled out as options for us: as things we can possibly choose to do. Although events clearly

[12] Ferrero, 'Constitutivism and the Inescapability of Agency,' 308.

happen which we might loosely speaking label immoral actions, the Kantian constitutivist view is that in a strict sense, the concept of an immoral action is incoherent.

This distinction between strict and loose ways of speaking will come up later, where I will say that it is clearly found in Plato's *Republic* (3.3.1). A point worth pressing now is that the kind of incoherence immoral actions are supposed to have is logical. As Allan Gibbard labelled her, Korsgaard in particular is a "moral logicist," if what is meant by this is "that morals are a part of the very logic of what to do."[13] When we are trying to act, what we are doing is trying to determine something to happen (3.1.2). However, we may have some apparent options which could not logically *be* things we determine to happen. In talking loosely of actions, we overlook that complication. We ignore what it would take for an action to be chosen. In talking strictly, we follow the logic, and discount as actions any which are incoherent. On a view like Korsgaard's, immoral actions are exactly incoherent, impossible packages. Being impossible packages, immoral actions are not things which can be done.

In the *Sources*, Korsgaard suggests that the answer to the question why be moral "must be one that makes it impossible, unnecessary, or incoherent to ask why again" (SN 33). In a sense, on her view, we even make a mistake in asking for a reason not to be immoral. The sense is that immorality is not one of the options, and so not one of the options which there might be reasons for. If being immoral just meant disobeying some authority – like some normative force on a realist view (1.2.1) – then we could ask why we ought not to do so, since disobeying that authority is an option open to us. If Kantian constitutivists are right, however, then being

[13] Gibbard, 'Morality as Consistency in Living,' 143. Velleman, on the other hand, is not a moral logicist (2.1.4).

immoral is disanalogous to disobeying an authority. In a strict sense, it has to be a kind of failing to do anything.

2.1.4

This second component, Possibility, and the first one, Agency, were the two which I suggested were shared by all kinds of constitutivism in ethics. We find a version of Possibility, for example, in the argument of David Velleman's *How We Get Along*, where he claims that:

> A person may feel categorically forbidden to do something because it is, as we say, unthinkable. To say that it's unthinkable means, in my view, that he cannot find anything in his self-conception to make it intelligible as something that he would do. Contemplating this ostensible option, the improviser finds himself at a loss to make sense of his doing it; and so, as an agent, he is at a loss as to how he could do it, that is, how he could make it his doing.[14]

As on Korsgaard's view, for Velleman, the impossibility of forbidden actions is in some way conceptual. The content of a particular apparent option clashes with how the agent conceives of their identity as one. The agent is unable to see how acting in some way could be something that they, the agent, do. This is important because, for Velleman, an action must strictly be something you can see yourself as doing. For the agent, or "the improviser," then, unthinkable actions are not open to them. Strictly speaking, logically or conceptually, unthinkable actions are impossible. As such, they are not options for the agent in their trying to decide what to do.

[14] Velleman, *How We Get Along*, 108.

For Velleman, the pressure to make sense to yourself leads in "pro-moral" directions.[15] On his view, "practical reason is not itself moral," but what it does demand "has encouraged us to develop a moral way of life." This is a different view to Korsgaard's: although Possibility and Agency are components of both views, for Korsgaard, the connection between agency and morality is more immediate. For Velleman, what agency immediately connects to is not morality but self-understanding. Our actions must be ones which we are intelligible to ourselves in doing. It just so happens that the process of trying to be intelligible to ourselves has led us in moral directions – and directions plural, not a single one.

2.2 How we know morality applies to and binds us

2.2.1

Those first and second components, Agency and Possibility, relate to *how* moral standards are applicable to us and binding on us respectively. Moral standards apply to us because we stand in some particular relation to the problem of agency. They bind us because they describe at least some of the conditions of our solving or engaging with that problem. The third and fourth components relate to how we *know* that moral standards bind us and apply to us as they do. Those components I said I would refer to as "Practicality" and "Indefeasibility."

2.2.2

The sense in which Kantian constitutivism is practical is that your knowledge of moral standards is knowledge of what you yourself are proposing to do in your attempts at action. As Korsgaard puts it, it is a kind of "self-knowledge … not knowledge which is then to be applied" (SC 67). It is not knowledge of something which already exists, which we are trying to describe. The kind of self-knowledge Korsgaard means is not knowledge of facts like our true

[15] Velleman, *How We Get Along*, 2.

motivations for acting, but rather of what we see ourselves as trying to do and what is logically entailed by that.

In this way, knowledge of moral standards is supposed to be available *a priori*, without experience. We can discover it by introspecting what we see ourselves as trying to do in acting. Introspection is not in general infallible, but that introspective starting point is a secure foundation for moral knowledge because what we *see* ourselves as trying to do is one half of what, in this sense, we *are* trying to do. The other, second half is just what the first half logically entails. Since we can miss logical entailments, it is this second half we can be mistaken about. Moreover, these first and second halves of what we are trying to do can contradict each other. This is what is claimed to happen when we try to act immorally: we try to be agents in ways that agents cannot.

In Kantian constitutivism, I am saying, moral standards belong to this second, entailed half of our attempts to act. From what we see ourselves as trying to do, a regressive or transcendental argument can show that it requires being moral. If what we were considering trying to do was immoral, we have discovered a contradiction, and so the action is impossible. It is bent in against itself, requiring us both to be immoral and to not.

A moment ago, I described knowledge of moral standards as knowledge of what you yourself are proposing to do in acting. Of course, we can also have some knowledge of the moral requirements which others are under. Kant is confident we know that all rational beings are bound by the same categorical imperative.[16] As I will go on to say, the reason why Kant can have this confidence is that it is part of the concept of a rational being for him that it has some common, minimal agential self-conception (3.1.6). The same premises which make the

[16] Rational in the sense Kant usually means that term: see 3.1.6 and its note 31.

Kantian moral argument work for us hold for any rational being by definition. A being for which those premises did not hold would not be a rational being. It does not follow from this in any way that *we* are rational beings, or that there are any anywhere. However, I will be arguing later on that we can recognise those premises do apply to us (3.1).

2.2.3

Continuing with Practicality, another way to frame it would be that knowledge of moral standards is knowledge of the nature of a problem which we see ourselves facing, as we see ourselves facing it. The relationship between these two framings can be seen in this way. Built into what you see yourself as proposing in doing something is what you see that proposing as being *for*: what problem you see yourself as engaging with. I believe that putting the idea in this second way makes it easier to see where it may have appeared before in the history of philosophy.

For Kantian constitutivists, we see ourselves as engaged with the problem of trying to choose our actions, and normative facts exist, so to speak, in the scope of our engagement with that problem as we conceive of it. As some have read Thomas Hobbes, for example, he holds a view which may seem similar. As Stephen Darwall reads him:

> Desire-and deliberation under its influence-are, Hobbes believes, what give ethical thought its point and function. So long as we are alive, we desire. And so long as we desire, we find ethical thought and, perforce, deliberation, unavoidable. Hobbes addresses *Leviathan*'s central normative claims, the laws of nature, to his readers as deliberating agents. As theorems about what leads to self-preservation, these provide

> lemmas that can be combined in each agent's practical reasoning with a normative thought each finds unavoidable under the influence of a desire he cannot shake.[17]

For Hobbes as Darwall reads him, we see ourselves as facing, inescapably, the problem of self-preservation. As we see ourselves facing it, the claims of the *Leviathan* are solutions to that problem. They are not metaphysical value claims, or expressions of Hobbes' acceptance of some norms or plans. They are factual claims about how to solve a problem which we might recognise ourselves as having to engage with.

Others have read Descartes' *Meditations* in a similar way. On such a reading, Descartes conceives of the problem of knowledge as the problem of how to hold beliefs which no further experience could shake. Descartes is not concerned with the problem of what is true or "real" as we usually understand it. He is concerned with the question of what to believe, and he starts with what he has to – he cannot help but – believe. For as Descartes sees it, some of our perceptions:

> are so transparently clear and at the same time so simple that we cannot ever think of them without believing them to be true. The fact that I exist so long as I am thinking, or that what is done cannot be undone, are examples of truths in respect of which we manifestly possess this kind of certainty. For we cannot doubt them unless we think of them; but we cannot think of them without at the same time believing they are true … Hence we cannot doubt them without at the same time believing they are true; that is, we can never doubt them. (AT 7:145–46)

[17] Darwall, 'Normativity and Projection in Hobbes's *Leviathan*,' 319.

Read this way, the *Meditations* are a historical example of Practicality because the way that Descartes makes progress is by looking to how we think of our problem of what to believe. The ideal belief is one which no circumstances would make as swap out for its contradiction. As it happens, there are some beliefs which we cannot swap out, in that "we can never doubt them." Those beliefs, then, are ideal beliefs: not true, but problem-solving ones. They are we ought to believe in the sense that they answer our problem what to believe. Descartes never sounds more Kantian than when he says that beyond this, "there are no further questions for us to ask: we have everything that we could reasonably want" (AT 144–45).

2.2.4

There is one more component to introduce before I explain how they fit together. In addition to Agency, Possibility, and Practicality, this fourth component is Indefeasibility. What it amounts to in the form it takes in Kantian constitutivism is that we cannot have knowledge which would defeat or undermine our having to grapple with the problem of agency.

Part of the Kantian argument, somewhat simplified, is that agency presupposes some kind of free will. We must have free will to be agents in that agents determine things to happen, and so must have the capacity to. Indefeasibility is then that we cannot have knowledge that we lack that kind of free will. Of course, it does not follow that we do know we have it, but the argument does not suppose that we do. What it supposes is that we have to assume we have that kind of free will in our trying to be agents. A certain practical, free-will-laden worldview is forced upon us for the purpose of the problem of agency. Here Indefeasibility interlocks with Possibility. There is no way of being an agent on which we do not have free will. When trying to be an agent, then, we can have no options inconsistent with our having free will. Since we have to try to be agents, we must assume we have some options: some ways of being

them. Therefore, we must assume we have free will, since otherwise we would not have any ways of being agents.

The presuppositions of agency – what is required to be able to determine things to happen – we might call "practical conditions," by analogy with what Allison calls Kant's "epistemic conditions." Allison defines an epistemic condition as "a necessary condition for the representation of objects, that is, a condition without which our representations would not relate to objects."[18] Roughly, epistemic conditions are the things which would have to be true for the world to be the kind of place we could know things about. It must be structured by space and time, for example. We need things to have that structure to experience them. Therefore, to know the world through our experiences, we need the world to be spatial and temporal.[19]

Where epistemic conditions are conditions of knowledge, practical conditions are conditions of agency. Indefeasibility, in my sense, is our inability to know we lack our practical conditions. What is entailed by this indefeasibility is that we cannot be faced with a conflict between the worldview forced upon us as agents, and the one we form in trying to understand the way the world is. In the third section of the *Groundwork*, we find Kant worrying that:

> with respect to the will, the freedom ascribed to it seems to be in contradiction with natural necessity; and at this parting of the ways reason *for speculative purposes* finds the road of natural necessity much more traveled and more usable than that of freedom; yet *for practical purposes* the footpath of freedom is the only one on which it is possible to make use of our reason in our conduct; hence it is just as impossible for

[18] Allison, *Kant's Transcendental Idealism*, 11.
[19] We need more than this, however – what we might call "theoretical conditions" – for the world to be not just a place we can know, but have a model of which enables us to act in it. See AR 32–35.

> the most subtle philosophy as for the most common human reason to argue freedom away. Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the very same human actions, for it cannot give up the concept of nature any more than that of freedom. (G 4:455–56)

If we could somehow know that we lacked free will, and that there therefore was no point in our trying to be agents, Kant seems to think that reason would be thrown into some kind of disarray. We might imagine this disarray being like the wasteland of the original preface to Kant's first *Critique*: a "battleground of … endless controversies" between dogma and scepticism (C1 Aviii). Indefeasibility would mean avoiding this. Our relationship to agency, and so to morality, would not be undermined by what we can know about the world as it "really" is.

2.2.5

Korsgaard gives what seem to be two different arguments for Indefeasibility. The first, recognisably Kantian argument is that if we had free will, it would not show up in experience. Freedom is an Idea in the Kantian sense, like God, immortality, and the political Idea "of the unity of a people *as such* under a powerful supreme will" (M 6:361). It is something "to which no object given in experience can be adequate."

Suppose we were as confident as we could be that all our behaviour is caused by the prior state of the world. Our best scientific theories posit laws which we see can predict all our choices before we make them. Kant does not see this as some scifi fantasy. He imagines having "such deep insight into a human being's cast of mind … [that] we could calculate [their] conduct for the future with as much certainty as a lunar or a social eclipse" (C2 5:99). Still, Kant insists, we "could nevertheless maintain that the human being's conduct is free."

The reason is, essentially, that we cannot perceive the causes of things. There is nothing contradictory about the idea that the whole chain we see as causes leading up to our choices might have actually been set off by those choices, in some timeless or timebending way. Kant's claim is not that we have any evidence for this being the case. Rather, it is that our inability to perceive causal relations means we could never have evidence either way.

Korsgaard does make this argument for the Indefeasibility component, but she also makes a more radical one. It turns on the fact that we are doing different things when trying to act in the world, approaching it practically, and when we are trying to "really" describe or model it, approaching it theoretically. In 'Creating the Kingdom of Ends' – the essay collected in the book – Korsgaard fits the two arguments together. She begins:

> As thinkers and choosers we must regard ourselves as active beings, even though we cannot *experience* ourselves as active beings, and so we place ourselves among the noumena, necessarily, whenever we think and act. According to this interpretation, the laws of the phenomenal world are laws that describe and explain our behavior. But the laws of the noumenal world are laws which are *addressed to us* as active beings; their business is not to describe and explain at all, but to govern what we do. Reason has two employments, theoretical and practical. We view ourselves as phenomena when we take on the theoretical task of describing and explaining our behavior; we view ourselves as noumena when our practical task is one of deciding what to do. The two standpoints cannot be mixed because these two enterprises – explanation and decision – are mutually exclusive. (CKE 204)

This suggests that, in principle, we could live with the kind of conflict between these standpoints that Kant was careful to argue does not occur. We could live with it because those

standpoints are not strictly contradictory: their claims talk past each other. How we must view the world when acting in it, and when modelling or explaining it, can in principle we different.

This latter, stronger argument raises more complicated questions than I can hope to deal with here. To cut to the heart of it: it seems to depend on a denial that "the truth" – some fact of the matter – is generally normative. As Allison reads Kant, this denial is at the heart of Kant's critical idealism. As I have argued elsewhere, I believe that Allison and Korsgaard share this understanding.[20]

2.3 How the four components are related

2.3.1

So far in this chapter, I have been explaining the four components into which we can break down Kantian constitutivism. The first was that agency is the site of moral standards. The second was that moral standards are conditions on the possibility of our being agents, as we are trying to be. The third was that our knowledge of moral standards is self-knowledge, of what we are proposing to do in trying to be agents. The fourth was that our need to try to do that as we conceive of it cannot be undermined by any knowledge we could have.

What I will be arguing now, in this section, might sound rather architectonic, to the extent that one properly can in breaking a system down rather than building it up. In Kant as elsewhere, however, architectonic can serve a useful purpose. One result it can make easier to see is when a system is complete, all pieces put together. Earlier (2.2.1), and in my headings, I gestured at one axis on which the components in this chapter split. Agency, the first, captures the type or kind of way that moral standards apply to us. Possibility, the second, captures the

[20] In my 'The missing two-thirds of metaethics.' See especially Allison, *Kant's Transcendental Idealism*, 45–59.

type or kind of way in which they bind us, given they apply to us. These are answers to the questions of *how* moral standards apply to us and bind us.

The other, latter two components match up to different kinds of questions: not of how, but of how we *know*. Practicality is how we know the Kantian constitutivist's moral standards bind us. They do so through conceptual impossibility, and we know about that impossibility because the concepts involved are our own in our trying to be agents. Indefeasibility is how we know that moral standards apply to us. The practical conditions are indefeasible, and being secure in our having to assume them, we know morality applies to us.

The other axis or division is right at the surface here. Agency and Indefeasibility relate to the applicability of moral standards: its kind or medium, and how we know about it. Possibility and Practicality – pair the Ps – instead concern the bindingness of moral standards. Again the former is the kind of bindingness, and the latter our access to knowledge about it.

2.3.2

What we have here, to be *very* Kantian, is a table with two columns and two rows. The columns are the hows and how-we-knows; the rows, applicability and bindingness. To illustrate:

	How	How we know
Applicability	Agency (1st) is the site of moral standards	Indefeasibility (4th) of our practical conditions
Bindingness	Possibility (2nd) is conditioned by moral standards	Practicality (3rd) of the concepts which rule out immorality

I said a moment ago that architectonic can make completeness easier to see, and the rows and columns here cover the general questions meant to be taken up by Kantian constitutivism. That first column, as I have said, captures the characteristics or aspirations of any constitutivism. Moral standards are normative because if we are not following them, we are failing – Possibility – to be agents – Agency.

How that claim gets a grip on us, or how its objects relate to us, is then answered on the right in a Kantian way. Agency is an Idea we cannot help but act under, and all it takes to know its content is reflection and logic. Part of the sense in which we are under it necessarily is that no knowledge we could have could warrant reason overthrowing it. The Kantian denial of metaphysical knowledge leaves room for the freedom required for agency to have "objective though only practical reality," as Kant puts it in the second *Critique* (C2 5:48).

2.4 Constitutivism and constructivism

2.4.1

This talk of practical reality offers a link back to two questions I set aside earlier: those of the relationship between constitutivism and constructivism, and of the distinctiveness of their metaethics (1.2.1). To take the former first, I claimed that, as I understand it, Kantian constitutivism is the form which Kantian constructivism takes in ethics. I can now be more specific about this. Kantian constructivism is in general Practicality. This was that knowledge of normative standards is knowledge of the nature of a problem which we see ourselves facing, as we see ourselves facing it. Accordingly, we find Korsgaard saying:

> according to constructivism, normative concepts are not (in the first instance …) the names of objects or of facts or of the components of facts that we encounter in the world. They are the names of the solutions of problems, problems to which we give

> names to mark them out as objects for practical thought. The role of the concept of the right, say, is to guide action: the role of the concept of the good might be to guide our choice among options … To produce a constructivist account of the right or the good is to ask: is there some feature of the problem itself, or of the function named by the concept, that will show us the way to its solution? (COA 322–23)

As problems pair with solutions, so concepts pair with conceptions. This terminology is Rawlsian, and Korsgaard uses Rawls as an example. She continues:

> The feature of the problem of liberal justice that shows us the way to its solution, according to Rawls, is that a liberal society must respect the freedom of its members while enabling them to pursue their conceptions of the good. The feature of the problem of free action that shows us the way to its solution, according to Kant, is that free action must be determined by the agent herself. (COA 323)

This "problem of free action" is the problem of moral philosophy: of what we ought to do. Practicality, I have been saying, marks Kantian constitutivism as constructivist. What we can see Agency as doing is confirming the problem it is constructivist about. Agency and Practicality together mark Kantian constitutivism as constructivism in ethics.

2.4.2

It has not been obvious to everyone that Kantian constructivism in ethics is possibilistic in the way that Kantian constitutivism clearly is. For example, Enoch notes that some have tried to read Korsgaard's *Sources of Normativity* as constitutivist. For his part, however, "the relation

between the views expressed" in Korsgaard's explicitly constitutivist works and in the earlier, constructivist *Sources* "is not entirely clear."[21]

The relation goes this way. To act is to solve the problem of having to choose how to act. To call something an action is to label it as a solution to that problem. These are constructivist claims, but note that it follows that a failure to solve the problem will fail to earn the label of an action, since actions are solutions to the problem. To work out what is constitutive of an action is to work out what it takes to count as one, and so is at the same time to work out what it takes to solve this problem. For some standards to solve a problem is for them to be constitutive of solutions to that problem. Constructivist theories how to act can – trivially – be framed as constitutivist ones.

The relation goes a particular way: constructivism can be framed as constitutivist. The core of constitutivism is just Agency and Possibility. I am suggesting now that a Kantian constructivism in ethics will involve Possibility. One could still be a constitutivist, however, without being committed to Practicality, the core of constructivism. Perhaps the nature of agency in the relevant sense is not something we can discover through reflection, but is only accessible empirically through bleeding-edge insights from behavioural science. Perhaps, alternatively, the reason we should care about being agents is more realistic and teleological than that it is what we are trying to be. These would be ways of being constitutivist in some sense without being Kantian constructivist. Practicality seems to *entail* Possibility, but Agency and Possibility do not entail Practicality.

[21] Enoch, "Agency, Shmagency," 171.

2.4.3

Enoch specifically mentions the *Sources*. There, Korsgaard's Kantian constructivism has an earlier name, "procedural realism," contrasted with substantive realism.[22] As Korsgaard explains that contrast:

> What distinguishes substantive from procedural realism is a view about the relationship between the answers to moral questions and our procedures for arriving at those answers. The procedural moral realist thinks that there are answers to moral questions *because* there are correct procedures for arriving at them. But the substantive moral realist thinks that there are correct procedures for answering moral questions *because* there are moral truths or facts which exist independently of those procedures, and which those procedures track. (SN 36–37)

A common misreading of what Korsgaard is saying in this and similar passages is that, on her view, it is a moral fact that there are right procedures which will lead us to right actions. Those who misread Korsgaard this way sometimes also take it to be her view that there being right procedures precedes there being moral facts. It ought to give us pause that this combination of ideas is incoherent. For any procedures to be right would be for there to be a moral fact or truth that those procedures were right. That fact, however, could not both precede those procedures and be a product of it. The prefix, "procedural," cannot just be modifying what Korsgaard is a realist about.

It is significant that nowhere in the *Sources* does Korsgaard call the procedures which lead to moral truths the "right" ones. She instead calls them the "correct" procedures, as in the

[22] Korsgaard identifies the two terms – procedural realism and constructivism – in, e.g., NG 389. In 'Realism and Constructivism,' she describes herself as "defer[ring] to prevailing usage" in calling her own view constructivist (COA 310).

passage quoted above. Those procedures are correct not in the sense that there is some moral fact which makes them normative for us, but in the sense that they reflect our understanding of the problem those procedures are designed to solve.

Consider the problem of agency. A correct understanding of that problem would be one which genuinely described the problem of having to choose our actions as we see it as facing us. A correct procedure would then be one which drew out of our understanding of that problem what it would take to solve it, showing us the standards we would need to be guided by in order to solve it. A moment ago I quoted Korsgaard as saying "the feature of the problem of free action that shows us the way to its solution, according to Kant, is that free action must be determined by the agent herself:" it must be autonomous (COA 323). The reason is not that autonomy is "right" in any substantive sense. Rather, the way we think of action, according to Kant, requires that it be freely chosen. To see ourselves as acting, then, we must be able to see our actions as freely chosen. That, Kant argues, requires autonomy, and so solving the problem of action requires autonomy. The correct procedure for that problem is being guided by what it takes to be autonomous. It is not the "right" procedure in any general sense: just the one it takes to solve the problem.

The important part of the passage I quoted from the *Sources* is the end, where Korsgaard says that substantive realists believe moral facts "exist independently" of any procedures. To rephrase what I have just been saying, the claim which Korsgaard is making here is not that the rightness of right actions is derivative of that of right procedures. Even correct procedures do not make actions right. The sense in which "there are answers to moral questions *because* there are correct procedures" is that there are correct procedures whenever problems are solvable: that is, when there are answers.

What makes something right, for a procedural realist or constructivist, is that it solves a particular problem. But accordingly, the rightness is particular: it cannot be detached from the problem being solved. What the prefix, procedural, changes about procedural realism is not its object but its scope. Procedural realists believe in normative facts, but only from the standpoints of the problems which they solve.

2.4.4

It may not be obvious that there is no realist way of accommodating this claim. I have not yet explained why there is not, but it is the question I am turning to. Firstly, I need to go back to the one component which I have not mentioned yet in this section: Indefeasibility, or our way of knowing moral standards apply to us.

Ordinarily, when you find you cannot solve some problem, you can give up on it: you can accept you cannot solve it, and move on to doing something else. That is not the case with your practical problem, the recurring one of having to try to choose what to do. In the senses I will break down later, that problem seems distinctively inescapable (3.2). Kantian constitutivists think that means that you can take it for granted that you are able to solve that problem. Since you have to try to choose what to do, you can take it as given that you are able to. In introducing Indefeasibility, I explained that Kantian constitutivists believe that there are particular conditions to your being able to choose what to do (2.2.4).

They believe, that is, that there are things which have to be true for you to have the capacity or the freedom to choose what you do. Those required truths, or practical conditions, are not claims we know in general to be true, but are claims which are given to us as true for the sake of trying to choose what we do. In that sense, they have a scope, or are internal to our practical problems. For the sake of trying to choose our actions, they are real facts, in the

Kantian sense of their being objective. They have no independent existence, however, in the sense they are not granted to us generally. Earlier I noted Kant's use of the phrase "objective though only practical reality" (C2 5:48).

Kantian constructivism is sometimes described as an alternative to both realism and anti-realism in metaethics, and the distinction which it makes between scopes or standpoints from which facts are given is key to this. The Kantian moral argument, as made in Kantian constitutivism, takes our practical conditions as premises. The soundness of that argument depends in that way on the practical conditions being given. Accordingly, however, that moral argument loses its force and its basis when it is considered from outside of the scope of our practical problems, from some other perspective.

Suppose, then, that we look at a Kantian constitutivist moral agent, and consider what they are doing. We can see their moral deliberations depend on an argument which is premised on the practical conditions. Seen from the practical perspective from which those conditions are given, what that moral agent seems to be doing is tracking a piece of reality: some form of realism seems to be true. The agent has some facts about the world and themselves, and from them they have managed to derive conclusions about what is possible for them, and so what they can consider doing.

Suppose, however, that we look at our example moral agent from some other perspective: one from which the practical conditions are not given, and so neither are the derivable truths. We can still see that the moral agent's deliberations depend on an argument from those conditions, but that argument can no longer be understood as an attempt to track some part of reality. From this other standpoint, the practical conditions are not given to the agent as real, and so the agent's following of the argument cannot be an attempt to find out more of

what is real. What the moral agent seems to be doing is simply acting out their own noncognitive acceptance of the practical conditions as being sources of norms for them.

2.4.5

What the moral agent is thinking from any perspective is: "I ought to do this." When we come out of the practical perspective, it does not change to: "Were the practical conditions true, I ought to do this." That would be to think a different thought, and one which was not immediately normative. That second thought only says that *if* something were true, then something would be normative.

Realism and noncognitivism are, at least in part, accounts of the relation between a moral agent's thought and the claims available to them. When the claims available change, the agent does not change: what changes is that relation. That may, as I am saying, have consequences for the proper account of that relation. As Korsgaard makes this point:

> considered in one way, constructivism and realism are perfectly compatible. If constructivism is true, then normative concepts may after all be taken to refer to certain complex facts about the solutions to practical problems faced by self-conscious rational beings. Of course it is only viewed from the perspective of those who actually *face* those problems in question that these truths will appear normative. Viewed from outside of that perspective, those who utter these truths will appear to be simply expressing their values. (COA 325)

This last point is elaborated in a footnote:

> expressivism, our latter-day form of non-cognitivism … is like realism also true after all, and also in a way that makes it boring. From the descriptive and explanatory

> perspective that is appropriate to scientific or perhaps in this case social-scientific inquiry, those who use normative language will appear to be simply expressing their values. When you are not in the grip of practical problems that provide standards for their own solutions, the truth and falsehood of statements employing concepts that embody those problems must be elusive. (COA 325)

It has seemed almost obvious to some that there is no distinctive space in metaethics for Kantian constructivism: that it must either be noncognitivist, as Allan Gibbard thought,[23] or if not error theoretic, then a form of realism. Enoch, for example, suggests the space of metaethics can be divided by "rather simple yes–no questions," and that when "the details of this little exercise are filled in … no room remains for a distinctively constructivist position."[24] Nadeem Hussain and Nishi Shah have thought that Korsgaard fails to distinguish her position from either realism or noncognitivism, instead apparently radically misunderstanding what metaethics is about.[25]

What is involved in Practicality or constructivism, however, is our having perspectives on the world which are correct for different purposes, and from which different claims are granted. Earlier I effectively suggested that Hobbes and Descartes could be read as early constructivists (2.2.3). For Hobbes, what would be granted in our deliberative problem is that we aim at our self-preservation. For Descartes, in our theoretical problem, it would be that we believe what strikes us, irresistibly. For both, those claims would then condition our possibilities. Inside of the perspectives of the problems they were working with, those claims would be part of our reality. If they turned those perspectives on metanormative questions,

[23] In e.g. Gibbard, 'Morality as Consistency in Living.'
[24] Enoch, 'Can there be a global, interesting, coherent constructivism,' 329.
[25] See Hussain and Shah's 2006 'Misunderstanding Metaethics' and their 2013 'Meta-ethics and its discontents.'

then they would appear to be realists. If they stepped out of those perspectives, however, and answered those questions, they would not need to be.

What is assumed in denying distinctive – or any – metaethical space for constructivism is that metaethical questions can only be asked from one perspective. If we take seriously multiple perspectives, however, then those questions can be answered variously. There is a question that comes before the standard flowchart or tree diagram: namely, that of from what perspective we are supposed to be answering.

2.5 Kantian constitutivism in review

2.5.1

If have just been explaining why Kantian constitutivists, or in general constructivists, see their views as being metaethically or metanormatively distinctive. They distinguish between standpoints from which the premises of problems are and are not granted. From the former, in our grappling with those problems, the relation in which we stand to their solutions is realistic. From the latter, in our grappling with different problems, the relation may instead be noncognitivistic.

Before then, I took up the question of the relation between constitutivism and constructivism. I suggested that Practicality, the core of constructivism, entails Possibility. Agency would then mark a constructivism as being in ethics, and so on these definitions, any constructivism in ethics could be framed as a form of constitutivism.

2.5.2

In the first two titled sections of this chapter, I introduced the four components which I see as making Kantian constitutivism. The first, Agency, was that agency is the site or the locus of

moral standards. The second, Possibility, was that moral standards are conditions on the possibility of our being agents, as we are inevitably trying to be. The third, Practicality, was that our knowledge of moral standards is self-knowledge, of what we are proposing to do in our trying to be agents. The fourth, Indefeasibility, was that our need to try to grapple with the problem of agency cannot be undermined by any knowledge we might have, such as that we cannot be agents.

I suggested where we might see those ideas as having appeared before or been foreshadowed in history. For Agency and Possibility, I pointed to Plato and the Stoics. I also compared how Possibility appears in Korsgaard's constitutivism and in the not-quite-Kantian constitutivism of David Velleman.[26] For Practicality, I considered ways of reading Descartes and Hobbes. For Indefeasibility, I looked to Kant, with his denial of metaphysical knowledge.

2.5.3

Then, in the middle of this chapter, I looked at how the four ideas I outlined fit together. Agency and Practicality are attempts to answer the questions of how morality applies to and binds us. Practicality and Indefeasibility instead address the questions of how we know that it does. The second division, I said, was at the surface here. Agency and Indefeasibility relate to morality's applying to us, rather than the way in which it is binding. That form of bindingness is instead what we get from Possibility and Practicality. Moral standards condition what we can do, as we understand what we are doing. I suggested that, together, the four components are how Kantian constitutivism answers the questions it is supposed to address, about normativity and the grip it has on us.

[26] In *How We Get Along*, 149, Velleman describes himself as having "a Kinda Kantian strategy."

3 In what sense do we face the problem of agency?

3.1 Rational and nonrational agency

3.1.1

The topic of this chapter is the problem which constitutivism theorises about or is addressed to. This is the problem of agency, and I will begin by covering what Kantian constitutivists mean by terms like "action" and "agency." This will quickly lead to a discussion of rationality, of problem-facing talk, and of what I believe are three distinguishable ways in which agency is inescapable.

Importantly, the claim that agency is inescapable does not mean we cannot fail to exercise or have it. We can fail to act by failing to make things happen, and of course, it will be argued, by trying to be immoral. In our ordinary talk, we do still describe self-conscious bad behaviours as actions. To make the claim they fail to be actions seem less unconventional, I will try to put it into some historical context.

In both Plato and Kant, we find distinctions between strict and loose ways of speaking. In introducing Possibility earlier, I said that in the loose way of speaking, we ignore what it would take for the object of an action to be something we were able to determine (2.1.3).

3.1.2

In the Kantian constitutivist's dictionary, trying to act is the way that rational creatures try to determine what happens – to their bodies, their minds, the world around them. To accommodate talk of nonrational creatures acting, we might distinguish between nonrational

and rational kinds of agency. When constitutivists of all kinds talk of agency, however, they typically mean the rational kind specifically.

In the Kantian sense, a rational creature is not simply or necessarily an intelligent one. That other sense of rationality – actually doing what rational principles require – comes closer to what Kant calls "wisdom" (Anth 7:200). But wisdom, Kant says, "is no doubt too much to demand of human beings." It is too much to expect "reason in [us,]" as he puts it in the *Groundwork*, to be "practical without hindrance" (G 4:449).

What defines a rational creature is that, in their own mind or from their own perspective, they find themself faced with a radically open version of the problem of trying to choose how to act. The openness of that problem can be brought out by a contrast. Suppose that our ends – our aims or purposes – were forced upon us, in the sense we did not ever have to think about pursuing them. They were just given and unquestionable, likes Descartes' clear and distinct perceptions: it was inconceivable that we might doubt them (2.2.3). All we had to do, and all that we could think of doing, was working out what acts or means best fitted with those ends.

This is not the kind of problem we find ourselves faced with in having to choose our actions. We do not find our ends or values or principles forced upon us in the sense of being deliberatively unquestionable. There may be only some ends which seem appealing to us, but we still have to choose to pursue them in our actions. Implicitly choosing ends for ourselves is part of what we do in choosing our actions.

Forget whether you think that we can "really" choose our ends or means, in the sense of whether our psychology "really" sets them. Thinking that the ends or means you will choose are the ones you were always going to is compatible with facing the problem of choosing them. It is compatible with facing that problem in that however sure we are that our actions are

determined for us, we still find ourselves having to grapple in our own minds with the deliberative task of choosing them. As Korsgaard puts it, you may believe that this grappling "is a sham, but it makes no difference" (CKE 163). Your having to try to choose how to act does not go away if you hold the right beliefs. As Thomas Hill has it, the claim is not "that those who sincerely believe in thorough-going determinism cannot act or, conversely, that no one who acts ever believes sincerely in thorough-going determinism."[27]

Not finding our ends or values or principles forced upon us is the sense in which I meant that a rational creature's practical problem is radically open for them. To put the point less carefully, in a deliberative respect, rational beings see themselves as having the *space* to call into question their bases for choosing actions as well as any candidate actions. As a matter of psychology, they may seem to have no hope of upending their natural principles or ends. The Kantian claim that a being is rational, however, is not strictly a claim about their capacities in that sense. The way in which capacities become involved is that a rational being may see itself as required to have some in order to rise to the problem of trying to choose its actions, as it cannot help but try to do (2.2.4).

3.1.3

The claim that we are rational in this sense is the most basic premise of the Kantian constitutivist argument. In an effort to be clear, then, let me risk being repetitive about it. A rational being is not one which *can* choose its own ends, as well as how to pursue them. It is not even one which *can* bring its ends up in reflection and entertain, perhaps without hope, its not pursuing them. The claim a being is rational is not essentially or immediately a claim

[27] Hill, *Dignity and Practical Reason*, 117.

about its capacities. It therefore is not falsified by the behaviour of the creature being predictable or in some sense predetermined.[28]

The claim a being is rational is a claim about the problems which a being has to deliberatively grapple with. That is, it is a claim about what a being finds itself confronted with having to work out. The phenomenology is not: I feel that I can choose my actions, which is a feeling that I have some kind of freedom. It is: I cannot help but be engaged in trying to work out how to act, as I would have to had I freedom.

There is a difference between sensing you have a choice, and experiencing yourself as having to try to make some choice. The former, different kind of claim is found in the work of libertarians about free will. In their paper 'The Phenomenology of Free Will,' Eddy Nahmias and colleagues collect some examples:

> For instance, the libertarian C.A. Campbell writes: 'Everyone must make the introspective experiment for himself: but I may perhaps venture to report … that I cannot help believing that it lies with me here and now, quite absolutely, which of two genuinely open possibilities I adopt.' Keith Lehrer says that such an experience 'accurately describes what I find by introspecting, and I cannot believe that others do not find the same.' And John Searle asks his readers to 'reflect very carefully on the character of the experiences you have as you engage in normal, everyday human actions' and tells them, 'You will sense the possibility of alternative courses of action built into these experiences . . . that we could be doing something else right here and now, that is, all other conditions remaining the same. This, I submit, is the source of

[28] If it *were* predetermined in the *right* way, that might show the creature could not be an agent. The Indefeasibility claim involves denying we can know or have evidence that it is predetermined in such a way (2.2.4).

> our own unshakable conviction of our own free will.' None of these philosophers concludes from these experiences that we in fact *have* an unconditional ability to do otherwise, but they do suggest that if we *don't* have such an ability, free will is an illusion. Our *experience* would be illusory.[29]

Our experience of being faced with having to try to choose our actions, by comparison, cannot be illusory. The psychological event, or the experience, is the being faced with that problem of choosing. It does not represent some fact beyond itself to which we might be unreliably related. It is the experience of having to try to fill a gap we do not also need to see as "really" existing.

3.1.4

I have been describing our practical problem – of choosing our actions – as *facing* or *confronting* us. This way of talking is common in the literature, but may be more obscure than is necessary. A way we might translate or ground this talk, I believe, is by turning it into being about our trying to solve some problems naturally or automatically. There may be certain problems which we catch ourselves engaged with, rather than having to take up self-consciously. The experience of being faced with or confronted with such a problem would be that of finding you were already trying to solve it. This is one of three things I will be going on to argue is meant in saying that agency is inescapable as a problem (3.2.2). We face that problem in that we are engaged with it without having chosen to become engaged with it.

[29] Nahmias et al., 'The Phenomenology of Free Will,' 165–66. I have omitted their added italics and underlining. The embedded sources are: Campbell, 'Is 'Freewill' a Pseudo-Problem,' 463; Lehrer, 'Can We Know That We Have Free Will by Introspection,' 150; and Searle, *Minds, Brains, and Science*, 95.

Consider an example of some problem-facing talk in Korsgaard. In 'Motivation, Metaphysics, and the Value of the Self,' Korsgaard responds to a critique by Paul Guyer:

> Guyer thinks what I mean by an argument made from the practical point of view is "an exploration of the consequences of our *choosing* to conceive ourselves as rational beings." But I don't mean that, for I don't think we have a choice about that. Our plight as self-conscious beings is that we find ourselves faced with the necessity of making choices and so in need of reasons to act. Kant's argument, as I understand it, is intended to start from that plight, to proceed from the standpoint of a being faced with the need for reasons and principles. (MMVS 62)

Korsgaard claims here that we are "faced with the necessity of making choices." Making choices is what we do at the end of a process describable as trying to make them. Either that process is up to us to begin, so to speak manually, or it begins automatically. If it were up to us to begin, then we would have the option of not beginning it. That being so, its end result would not be necessary. Korsgaard says it is necessary, and so what she has to mean in passages like this is that the process of making choices starts for us naturally, and we find ourselves caught up in it.

3.1.5

Suppose I have this right: that the experience of being faced with a problem is that of finding you are already trying to solve it. We might make this clearer still by explaining what is involved in, or meant by, that finding and that trying. Finding you are trying to solve a problem would seem to mean that you see your mind as making proposals with a certain conceptual content: that is, as being proposals of a certain type or kind. I suggested earlier that we can have secure knowledge of what our minds are proposing to us, since what they are

proposing in this sense is what we see them as proposing and what is logically entailed by that (2.2.2). It is the Kantian version of what Aquinas called "practical knowledge," or Anscombe "knowledge without observation." The reason why this is the right sense of proposing or trying is that it is the one which will lead to normative truths being backed by endeavours with which we identify.

To explain: suppose that what we meant by "finding you are already trying" to do something was that behavioural science told us that our psychology was in a clear sense aimed at something. It was aimed at something in the sense that we were always going to do what, quirks aside, our brains had calculated was the most effective way of pursuing that aim. That aim might be continuing the species, say. All our conscious mental activity could be convincingly explained as rationalising ways of inevitably, by default subconsciously pursuing that aim.

If we discovered we were trying to do something in *that* way, then it would not be normatively interesting. There would be no interesting sense of "ought" in which it followed that we ought to pursue continuing the species. That aim would still be one which we would see ourselves as able to take or leave at the deliberative level of trying to choose our actions. All that we would know is that there was a certain inevitability to those deliberations.

For the claim that we are trying to do something to be interesting, then it needs to be something we identify with, or something we are logically committed to as the way of pursuing some objective we identify with. That identification is what we have with things we see ourselves as trying to do, first-personally or reflectively. If you see yourself as trying to climb a mountain, and you discover that the only way is via some pass, then that fact is normatively interesting as long as you are trying to climb that mountain.

The interesting sense of finding yourself faced with some problem, I am saying, is discovering you are trying to solve it, in the sense you see yourself as making proposals in your mind which have the form of answers to that problem. To say we face the problem of acting is to say that we catch ourselves engaged in trying to act, where that involves proposing ways of acting to ourselves: ways of trying to determine what happens. It is not to say that we sense that we are able to act, if we mean something empirical by that. It is a claim about how a being sees its own aims and mental life.

3.1.6

The aim of this argument was to pin down the Kantian constitutivist sense of being rational. I claimed that in this sense, to be rational is to see yourself as faced with an open practical problem. That problem is open in that it is not given to you, deliberatively, that your actions will embody certain ends. I have just been claiming that being faced with the problem involves seeing that you are already, naturally trying to solve it.

One way to put some of this is that whether a creature is rational is, for Kantian constitutivists, a question of its self-conception. It is a matter of how it sees itself, not immediately of what for it is mentally possible. Admittedly, however, sometimes Kant and Korsgaard talk in ways that are stronger-sounding. For example, in one place Korsgaard describes rationality as "the capacity to ask whether something that would potentially motivate you to perform a certain action is really a *reason* for doing" it (FC 40).[30] Kant at one point calls it "a capacity to choose a way of life for [yourself] and not, as other animals, to be bound to a single one" (Conj 8:112).

[30] Claims like these could be taken to suggest that Korsgaard is committed to what I have elsewhere called an "empirical optimism" about our powers of reason. In my 'Kantian indifference about moral reason,' I gave reasons to believe that Korsgaard is only committed to what I called a "philosophical" optimism.

Claims like these might suggest that for Kantian constitutivists, being a rational creature comes down to one's potential, in the sense of one's abilities or the futures which are open. Strictly speaking, however, the sense in which for both Korsgaard and Kant we have rational capacities is that, to paraphrase the latter, we must act and think "*under the idea*" of our having them (G 4:448). First-personally, we see ourselves as faced with the choices which correspond to having rational capacities. For Kant, as I read him, the story here involves what he calls the "fact of reason" (C2 5:31–32). The fact of reason is Kant's name for what makes it seem to you that you are rational, not forced to unreflectively aim for whatever you believe that you desire. It does this by suggesting to you the possibility of taking actions on the grounds that you ought to take them, morally, rather than that you desire to. When you take that possibility on board, you come to see yourself as facing the rational practical problem in that you no longer see your principles or ends as deliberatively set for you.

Without the fact of reason, thinks Kant, "no amount of subtle reasoning" could have made us "even suspect the possibility" of being rational (Rel 6:26).[31] It is through it that Kant says that our capacity for rational agency comes to have "objective though only practical reality" (C2 5:48). That qualification – "practical" – is important, just as it was earlier for understanding constructivism (2.4.1). As I read him, the key to Kant's thought here is that we are entitled or warranted to take the problems which we find ourselves faced with inescapably to be ones which we can rise to solving. In the case of our rational version of the problem of trying to

[31] Kant muddies things in this footnote to the *Religion* by saying "the most rational being of this world … might apply the most rational reflection" and still not realise they could, in my sense, be rational (Rel 6:26). The kind of rational reflection Kant means, however, is only instrumental: in respect of set ends, "about what concerns their greatest sum as well as the means for attaining the goal determined through them." What I think Kant has in mind is that if we imagine a rational being and take away the fact of reason, then although their acuity or mental power, so to speak, will be unchanged, they have no hope of realising their rational potential.

choose how to act, that means that – only practically, inside the scope of trying to do that – we can take it for granted that we have the rational capacities required to do that.

3.1.7

For even the most intelligent of nonrational beings, in their minds their objectives must, so to speak, be set. This is not an empirical claim: it simply follows from what for Kantians it means to be a rational being. The reasoning or calculation of nonrational beings extends at most to working out what serves their objectives. That reasoning may be sophisticated, flexible, demanding – even things like altruistic. It will not count as rational, however, because of how its author sees, or fails to see, their principles and ends. To the extent that any nonrational creature is aware of their ends as such, those ends must seem innately self-expressive.

As I said at the beginning of this section, constitutivists of all kinds are typically concerned with only rational agency. Trying to act, for Kantian constitutivists, is rational beings' way of trying to determine things to happen. Rational agency differs from nonrational in that, practically, we see ourselves as able to question even our bases for acting, whether or not we often consciously engage in such questioning.

3.2 Practical inescapability

3.2.1

A little carefully, we might say that trying to act is our way of trying to do things. What need treating carefully are those cases in which we ordinarily say that we try to do nothing. For trying to act and trying to do things to line up, trying to do nothing needs to mean trying to do *something*. We might think of it as trying to keep your body still and silent, all else being equal – if nothing else is acting on it. When you try to do nothing, you try to determine that stillness and silence to be what happens, or that what happens to your body is whatever would happen

if only outside forces were acting on it. Since trying to act, for Kantian constitutivists, is a label for our way of trying to determine what happens, as some kind of attempted determining, trying to do nothing must be counted as a kind of trying to act.

Seeing this point is important to understanding how constitutivists can call agency as a problem inescapable. But it is not the only hurdle: another is the distinctiveness of the sense in which agency is inescapable. As Korsgaard puts it at the opening of *Self-Constitution*:

> Philosophers like to distinguish between *logical* and *causal* necessity. But the necessity of action isn't either of those. There's no logical contradiction in the idea of a person not acting, at least on any particular occasion. … And although particular actions, or anyway particular movements, may have causes, the general necessity of action is not an event that is caused. I'm not talking about something that works *on* you, whether you know it or not, like a cause: I am talking about a necessity you are *faced* with. (SC 1)

Rational necessity may be a kind which we are faced with, but what rational necessities say is that we must do or think things if we believe certain other things. The necessity of trying to act is not like that: we cannot get out of trying to choose what to do by giving up some premise or assumption from which it follows rationally that we have to try to choose.

3.2.2

A common example of how we can escape trying to act is by falling asleep. It is only when you are conscious, or more something like present, that you face the problem and the pressure of having to try to decide for yourself what you are going to do. Examples like sleep are worth considering, because although falling asleep is something which can happen to you, to *put yourself* to sleep you have to do something, like rest in bed or down some sleeping pills. Since

you have to do something for sleep to be a circumstance which you put yourself into, trying to put yourself to sleep is not, in *this* moment, a way of escaping the problem of agency. You try to act in this moment in order than in a future one, you might not need to do so.

This is one of at least three distinguishable ways in which agency is inescapable, which Korsgaard runs together as it being "our plight" and which I will call its "practical" inescapability.[32] That first way, just mentioned, is that we cannot determine our way out of trying to act in any moment. Once you are engaged with the problem of trying to choose an action, the only ways you can escape from that problem are by acting, by being passively distracted, or by something stunning you or knocking you out.

The second way that agency is inescapable, which follows from what Kantian constitutivists mean by trying to act, is that we have no ways of trying to determine what happens besides for trying to act. By the Kantian constitutivist definition, acting is whatever our ways are of determining what happens. If we have any way of determining what happens, then it counts as a way of acting. As Ferrero puts it, "to engage in any ordinary enterprise is *ipso facto* to engage in the enterprise of agency."[33] Agency, as he also says, "is the enterprise with the largest jurisdiction." Importantly, this means that if Kantian constitutivists can identify moral principles which bear on acting, then they bear on all the ways we have of determining things to happen.

[32] Korsgaard calls agency, or trying to act, "our plight" in a range of places. I gave one as an example above, in explaining talk of problem-facing (3.1.4; MMVS 62). In *Creating the Kingdom of Ends*, Korsgaard says that "choice is our plight, our inescapable fate, as rational beings" (CKE xi), and that "the basic problem [of moral philosophy], set by the plight of rational agency, is 'what should I do?'" (CKE xii). In *Self-Constitution*, she says that "the necessity of choosing and acting … is our *plight*: the simple inexorable fact of the human condition" (SC 2).

[33] Ferrero, 'Constitutivism and the Inescapability of Agency,' 308.

The third of the three components of agency's practical inescapability is the following: that when we find ourselves engaged with the problem of agency, it is not because we chose to be engaged with it. Trying to decide what to do is not an activity we choose to initiate, but is rather one we catch ourselves engaged in, so to speak automatically or naturally (3.1.4). Trying to initiate that engagement voluntarily would take trying to determine something to happen: namely, that you began the process of thinking about choosing your next action. In other words, in order to choose to start to try to act, you would need to be trying to act already, since that choosing to start to try to act would be a kind of mental action.

3.2.3

As I have been saying, practical inescapability is not logical, causal, or rational necessity. It is, firstly, not being able to determine our way out of trying to act in any moment; secondly, having no way of trying to determine what happens besides for trying to act; and thirdly, not needing to try to determine ourselves to begin the process of trying to act. What constitutivists believe is practically inescapable is agency as a problem: that is, trying to act, or trying to choose for ourselves what to do. At least in a strict sense, *having* agency, or *acting*, are not themselves inescapable. We can fail to act, or fail to be agents: the "trying" is ineliminable here. In other words, it is the engagement with the problem of agency which is inescapable, and not our success at it.

A familiar way that we can fail to act is by failing to bring about whatever event, or property of an event, we were trying to bring about. You try to climb the stairs, but you trip, and instead you end up falling down them. What you tried to determine to happen did not, and so you failed in that attempt at acting. More abstractly: I try to win an essay contest, but you enter too, and the judges pick your entry. I tried to determine it to be the case that I won, but I failed, and so on that scale I failed in my acting. I succeeded at writing my essay and at

entering the competition, but I was trying to make more than that happen: I was also trying to win.

3.2.4

Aside from such failures to be efficacious – "in bringing about whatever state of affairs [you] intended to bring about through [your] action" (SC 82) – Kantian constitutivists may so far seem to have an extremely inclusive understanding of acting. That understanding includes doing nothing; mental acts like starting thinking; and both entering contests and winning them, as distinguishable attempts at doing something.

A commonly claimed difficulty for constitutivists in ethics is that of "developing a conception of action that is minimal enough to be independently plausible, but substantial enough to yield" any pro-moral content.[34] As we will see, Kantian constitutivists like Korsgaard do believe they can derive moral content from the minimal conception of action which I have so far been sketching (5.1). I will be arguing before then that this is in part because that minimal conception can be seen to give rise to a sophisticated, recognisably Kantian philosophy of action (4.2–4.3).

Our point for now, however, is that if some moral content really could be found in action or agency, then for constitutivists like Korsgaard, it would provide another possible way of failing to act. Kantian constitutivists try to argue that immoral actions are in some way failures as actions. When you try to act, but what you attempt in particular is some immoral package, constitutivists seek to argue that in some strict sense you fail: you do not manage to act.

[34] Katsafanas, *Agency and the Foundations of Ethics*, 2.

That talk of a strict sense is what I want to turn to now. Denying the status of actions to instances of immorality seems to fly in the face of ordinary language, where we call immoral actions that, exactly. This concern might be eased, however, by drawing a distinction between two ways of speaking. In this we can follow the lead of Thrasymachus, the first character in Plato's *Republic* to sound like a constitutivist.

3.3 Action loosely speaking

3.3.1

Halfway through Book I of the *Republic*, Plato has Thrasymachus accuse Socrates of "trying to catch people out as usual" by running together strict and loose ways of speaking about ruling or governing. Thrasymachus begins:

> I suppose if someone makes a mistake in treating sick people, you'll call him a doctor, just insofar as he makes that mistake? Or accountants – will you call someone who makes an adding mistake an account, at the moment that, and insofar as, he's making that mistake? I think we do talk like that; we do say "the doctor," "the accountant," or "the writing-teacher" made a mistake; but in fact, I imagine, none of these, insofar as he is what we call him, will ever go wrong. So, strictly speaking, since speaking strictly is what you go in for, no expect in any trade makes mistakes. It's when expertise has failed him – when he's *not* a craftsman – that he goes wrong. (Rep 340d–e)

Thrasymachus then returns to the topic of ruling. He continues:

> no craftsman, no expert, no ruler ever makes mistakes at the moment at which he *is* a ruler; even though in fact everyone would say "the doctor made a mistake," "the ruler made a mistake." Well, take me as having answered you just now in this other, loose

> kind of way; but actually the strictest way of putting it is to say that the ruler doesn't make mistakes, insofar as he's a ruler; that if he doesn't make mistakes, he lays down what is best for himself; and that this is what those he rules over need to do. Thus it's as I was saying all along: what it's just to do, I say, is what is in the interests of the stronger. (Rep 340e–341a)

For all of these constitutivist-sounding claims, that we must meet certain standards to count as crafting or ruling, Thrasymachus is not a constitutivist, and what he says here plays no real role in his argument. Socrates takes it to play one nevertheless, and this is part of why Thrasymachus gets frustrated, is hesitant about Socrates' replies (Rep 342d), and has to set them aside and give what Plato calls a speech (Rep 344d). What Thrasymachus appears to believe is that what is normative for a person is what serves their interests. What distinguishes a ruler is that they mould the world such that serving their interests is in others' interests. When Thrasymachus eventually asks: "how… am I supposed to persuade you? If you're not convinced by the things I was saying," he is struggling to see how to persuade someone to take an interest in what serves their interests (Rep 345b).

The distinction Thrasymachus makes between strict and loose ways of speaking is a useful one in any case. Ordinarily, loosely, immoral actions are describable as instances of acting. Constitutivists are not committed to the view that our ordinary talk about bad actions is incoherent; and in the other direction, that ordinary talk does not disprove any premise of constitutivism. For Kantian constitutivists, at least, their claims about actions are not grounded in ordinary language, but on what we conceive of ourselves as doing in our attempts to determine things to happen. What they argue is that the way we conceive of our attempts at action is inconsistent with our acting immorally. That claimed entailment is what gives rise to a second, strict and technical sense of action.

3.3.2

Kant rarely talks as Thrasymachus and Socrates do about contingent roles like those of the doctor, the accountant, and the writing-teacher. An exception is in the preface to the second *Critique*, where we find him distinguishing between something being permitted or forbidden and it being dutiful or contrary to duty. It is, Kant says:

> *forbidden* to an orator, as such, to forge new words or constructions; this is to some extent *permitted* to a poet; in neither case is there any thought of duty. For if anyone is willing to forfeit his reputation as an orator, no one can prevent him. (C2 5:11)

The word translated as "reputation" here – *Ruf* – has a range of meanings, all suggesting some kind of vocation. For example, the same word is the "call" when Kant talks about "drown[ing] out the troublesome call of reason" (Pro 4:381). In the context, Kant is clearly talking about what is required to be an orator. He is not merely interested in what it takes to be seen as one, to the extent those two things part.[35] To put Kant's example in Thrasymachus' terms: no orator, strictly speaking, ever invents new language unfamiliar to their audiences. Anyone who does so gives up on being an orator to that extent, since there is no way of combining being an orator with forging novel language. Of course, loosely speaking, we might simply call an orator who acts that way a poor one; but we can also talk in Thrasymachus' stricter way and say they forfeit their status as one.

What we do find clearly in Kant, more than strict and loose talk about contingent roles, is talk of different ways of seeing an agent which might be translatable into ways of speaking. Kant

[35] Perhaps a preferable translation of this *Ruf* in the second *Critique* might be "status." Your status can mean your reputation, but it can also mean your counting as something – like your citizenship status. If I am right about how to read this passage, then it would seem to be an example in Kant of what Korsgaard calls contingent practical identities: normative roles we can in principle give up (SN 100–02; SC 19–22).

distinguishes between two ways of looking at ourselves in respect of our actions: practically, as things we have to choose; and theoretically, as things to understand our having done.[36] We famously find a version of this distinction in the *Groundwork*, when Kant talks about "thinking of the human being in a different sense and relation when we call him free and … a part of nature" (G 4:456).

It takes a clearer form, however, in the *Metaphysics*. There, Kant says:

> we can indeed see that, although experience shows that the human being as a *sensible being* is able to choose *in opposition to* as well as *in conformity with* the law, his freedom as an *intelligible being* cannot be *defined* by this … We can also see that freedom can never be located in a rational subject's being able to choose in opposition to his (lawgiving) reason, even though experience proves often enough that this happens (though we still cannot comprehend how this is possible). – For it is one thing to accept a proposition (on the basis of experience) and another thing to make it the *expository principle* (of the concept of free choice) and the universal feature for distinguishing it (M 6:226)

Kant distinguishes in that first sentence between sensible and intelligible visions of ourselves as agents. Roughly, the intelligible vision is the one we have of ourselves in choosing what to try to determine to happen. That intelligible, practical perspective is the one to which the strict sense of acting belongs. From it, the only way in which Kant says it makes sense for us to act is following the moral law. It is from the sensible, theoretical perspective that we see

[36] Korsgaard sometimes glosses this practical/theoretical distinction as a first-personal/third-personal one. A confusion I mean to avoid in sticking to the former terms is about whether we can take both perspectives on ourselves. We can: to try to understand your actions as events is to view them theoretically or third-personally. Korsgaard particularly uses the personal terms in *The Sources of Normativity* (SN 14–17, 95–96, 124, 257); and glossed that way, the distinction is central to her essay 'The Normativity of Instrumental Reason' (COA 56–57).

ourselves as – loosely – acting or choosing "*in opposition to* as well as *in conformity with*" the moral law.

Similarly, in the second *Critique*, Kant links the morality of an end to "the possibility or impossibility of *willing* the action by which … [that end] would be made real" (C2 5:57). Actions with immoral ends are not ones you can will, and therefore cannot strictly be your actions at all. They must at best be things which happen to you, even if, per Kant's *Religion* (5.3.3) – what happens to you in the moment can still be understood as being your fault.

Later on in the passage just cited from the *Metaphysics*, Kant calls "the possibility of deviating from [morality] an inability," rather than "really" – *eigentlich*, or strictly speaking – "an ability" (M 6:227). It has been suggested that the word Kant uses for an inability – *Unvermögen* – might mean a worthless ability, one not worth having, rather than an inability strictly.[37] One problem with this reading is that it sits uneasily with the rest of the passage. Kant is saying that our apparent capacity to act immorally cannot ground how we think of our agency, since it might be just apparent: we cannot see things' causes, and so what really drives our behaviour. Another problem is that elsewhere, Kant does seem to use *-vermögen* to mean ability or faculty literally, as when he says that "all human insight is at an end as soon as we have arrived at basic powers or basic faculties" (C2 5:46–47).

3.3.3

Of course, the fact that distinctions between strictnesses of speaking have been made before in the history of philosophy does not mean that Kantian constitutivists have alighted on a normatively significant one in respect of rational action. What it does contribute to showing, I

[37] For example, by Allison in his 'Morality and Freedom: Kant's Reciprocity Thesis,' 420–23. According to Allison, Kant calls our capacity to act immorally an inability "presumably because it cannot be ascribed to perfectly rational beings."

hope, is that at least one of the ideas which goes into Kantian constitutivism is not as radical or unconventional – given an eye to history – as it might seem. The idea that some event or activity might strictly speaking fail to be an instance of its type is at least as old as Book I of the *Republic*, even if Thrasymachus' use of it is misleading.

Korsgaard reads Plato's Socrates as being himself a constitutivist.[38] Although I agree, on my own reading, this does not become clear until Socrates' argument for the soul's immortality. That argument, in Book X, might be summarised like this:

1. Injustice is the existentially threatening kind of badness of a soul (Rep 608e–9b); but
2. However great it is, the soul still exists (Rep 609c–9d); and
3. So, souls have immortality (Rep 609d–10c).

This argument is really just like "one of those puzzles" Socrates earlier says "people like to play with" (Rep 422e). The puzzle he had in mind at that earlier point, in Book IV, is that of when a city is not a city. It gets its interest from an ambiguity in whether the second "city" is meant loosely or meant strictly. Socrates' argument for the soul's immortality similarly depends on such a shift in meaning.

What Socrates is running together are what we might call strict and loose senses of having a soul. The sense in which the worst tyrant of the *Republic* still has a soul is that his soul's parts have the potential to be united under laws again. Because his agency is indistinguishable from that of his appetite, however, there is also a strict sense in which he has become just his appetite. His soul is nothing more than its parts, and in that strict sense, it no longer exists. Injustice is "the soul's specific badness" taking a soul in the strict sense (Rep 609c). The

[38] See especially SC 133–176 or COA 100–126.

specific or congenital badness of a soul in the loose sense would be what broke its parts or their connections. That badness of a loose soul the soul would not survive, and so Socrates' argument malfunctions.

Socrates' talk of the bad as what "ultimately breaks [something] down and destroys it," however, is clearly constitutivist (Rep 609a). Standards of goodness condition the possibility of something going forward as the kind of thing it is. To borrow Socrates' example, "the specific badness of a body, disease, dissolves and destroys it by reducing it to a state where it's not actually a body at all" (Rep 609c). Once we see that Possibility is clearly here in the *Republic*, it becomes more obvious that Plato's references to agency might be understood constitutivistically.

3.4 Constitutivism clarified

3.4.1

The subject of this chapter has been the problem of agency: what it is, how we face it, and in what ways we might escape it. Just now, my topic has been the way we ordinarily talk about action. I called that a loose way of speaking – not derogatively, but to distinguish it from a stricter, more technical sense of action which follows from arguments like Kantian constitutivists'. It follows from those arguments in that they would show that how we think of our actions as we take them rules out those actions coherently – that is, logically consistently – having immoral content.

I tried to make this distinction between strict and loose ways of speaking seem less unusual by pointing to it in Plato. In the *Republic*, Plato has Thrasymachus make such a distinction in Book I. I related my discussion of that distinction to Kant, and then in passing took up the

question of whether Socrates in the *Republic* is a constitutivist, since Thrasymachus, despite appearances, is not.

3.4.2

I began the chapter by defining trying to act, for Kantian constitutivists, as rational creatures' way of trying to determine what happens, or how things happen. I distinguished between rational and nonrational creatures, with rational creatures facing a practical problem which is radically open, in that their ends or their aims are not deliberatively forced upon them.

To claim a creature faces such a problem, I said, is not to make a claim about what they can do. I distinguished between sensing you have a choice and experiencing having to try to make one. The former experience would be that of perceiving or intuiting your having some capacity. The latter would be that of being confronted with the problem that capacity was needed to solve.

I went on to try to pin down this talk of being faced or confronted with a problem. I suggested what it means in this context is that you catch yourself naturally or automatically trying to solve the problem. You see yourself as making proposals to yourself which have the form of answers to that problem. In the practical case, you catch yourself considering proposals what to try to determine to happen.

3.4.3

Constitutivists often claim that agency is inescapable. What I said they do not mean by this is that our acting is logically, causally, or rationally necessary. I distinguished between three components of what I called its "practical" inescapability. The first was that we cannot determine our way out of trying to act in any moment. To do so would be to act, and so to fail at escaping doing so. The second was that we have no way of trying to determine what

happens besides for trying to act. The third, which I earlier identified with agency being a problem we are faced with, was that we do not need to try to determine ourselves to begin our trying to act.

For all that, we can still fail to act. We can do so by failing to be efficacious. When you try to act, you try to determine something happens, but that something can refuse to come about. The sense in which agency is inescapable is not that we are guaranteed to have it. It is that we are always trying to have it, with that being an endeavour at which we can fail.

We can fail to be agents in another way, according to constitutivists in ethics. We can fail to be agents by failing to be moral. Two arguments for that conclusion will come up in the next two chapters; but their main topics will be how incentives and ends get into the Kantian constitutivist philosophy of action.

4 Why should our moral psychology feature incentives?

4.1 Realism and constructivism

4.1.1

In the previous chapter, I claimed that what Kantian constitutivists mean by trying to act is trying to determine what happens: that particular events occur, or the way that they do so. I then moved on to their claim that we face the problem of trying to act understood that way. I tried to clarify that problem-facing talk, and explain what might be meant by calling the problem inescapable.

This chapter rewinds to trying to act being trying to determine what happens (3.1.2). It considers what claims about action, if any, can be seen to follow from that austere definition. I will be arguing that some claims do follow: I think the whole Kantian philosophy of action. I will be focusing in particular on the concept of an incentive, and on the package of an action including an end.

This will be a difficult argument, but not one I believe that Korsgaard anywhere makes. Nevertheless, Kantian constitutivists need to make it, in the following sense. The way that Kantian constructivist theories set themselves up is as being fundamentally premised just on how we conceive of the problem which is to be solved. What I mean by "fundamentally" is that the principles of the theory are meant to follow from this kind of narrow basis.

We might still need to bring more premises in to draw particular conclusions from those principles. For example, suppose Kantian constitutivists succeed in arguing for Kant's

categorical imperative. That imperative itself is just a principle: it needs to be applied to rule out anything specific. To draw specific conclusions, we need to feed some further content into that imperative. We need to know our options, people's ends, who around us we compete with for means: those kinds of further premises.

For Kantian constitutivism, the problem is the practical one of what to do. What the theory aims to show is that, surprisingly, some version of Kant's categorical imperative can be derived from only the nature of that problem as we find ourselves faced with it. If it works, the attraction of this approach is exactly the austerity of its premises. Being faced with a problem is not something we have to choose, the way that we would have to choose to take on further premises.

4.1.2

To illustrate the point here, consider a scenario. You are faced with the problem how to act. You have to choose what to do, but you find yourself struggling to see how to make that choice. A realist leans over, and tells you: "Do what maximises happiness!" You ask them why you ought to do that, and they tell you some story about presumptions of equality and our valuing of happiness. This story seems appealing, in its way, but it does not yet have you satisfied. You press the realist on what their premises are, or what they start by taking for granted. They confidently answer: "Obviously, happiness is valuable; every sane person accepts that." They accept it in their own case, at least, and an argument can get started with that. You value your happiness; you have no grounds for denying the same value to others' happiness; and so – to skip just a few steps – you are committed to the value of all happiness.

This argument speaks to you, and you are minded to accept it. You can work out the logic behind its steps, so all you now need to do is accept its premise. Happiness does seem

obviously valuable – but then into your mind creeps the realisation that whether to accept the premise is a problem, which you are now engaging with. What this realism was supposed to do is solve for you the problem how to choose to act. The way it seems to have done so is by replacing that problem with another, namely whether to make some judgment.[39] Now you need a theory to help you solve that other problem instead. Realism does not answer problems: it passes on the task of answering them to someone else.

This is not meant as some kind of ad hominem, of course. It is only a dramatisation of the argument I briefly gave in my introduction for realist theories being necessarily inadequate (1.2.1). Realism cannot ultimately answer the question of why we ought to do things. It either has to end in a claim we *just do*, or not give the ultimate answer itself.

4.1.3

A constructivist leans over: "You were asking how to act?" You confirm, and they start asking what you mean by that. What you mean, it turns out, is that you find you have to try to determine how or that things happen. The constructivist seems interested in two things. The first is that you cannot get out of trying to solve this problem. They argue that for the purpose of that problem, you can therefore take for granted the conditions of your solving it (2.2.4). The second is this idea of determination. They begin to try to show that you can get somewhere by working out what it would take for you to be able to determine anything.

The constructivist does not ask you to accept any premise. What their premises are, in the logical sense, are the features of the problem they are trying to help you with. They may seem to be trying to do the impossible – to draw meaningful conclusions from those features – but

[39] I do not have space here to argue that the methods of realism will not work any better for judgments like this. For Korsgaard on something like that issue, see her 'The Activity of Reason' (AR).

whether they succeed or not is another matter to the one I am attempting to bring out here. The constructivist, if they succeed in being one, does not replace your problem with a different one. They do not ask you to accept anything which is not already built into the terms of your problem.

This is the attraction of Kantian constructivism: it promises to solve problems, not pass them on. For Kantian constitutivism to have that attraction, it needs to build up from the features of our practical problems. But the only obvious feature essential to our practical problems is that what we see ourselves as trying to do is trying to determine things to happen.

4.2 The demands of our practical problems

4.2.1

In part of her reply to Nagel in *The Sources of Normativity*, Korsgaard describes in the following way how she believes we think of causes:

> Our ordinary notion of causality seems to combine two quite different ideas: the idea of *power* or (there is no non-redundant way to put this) of one thing *effecting* another or *making* another happen, and the idea of universality – that this occurs in a regular or law-like way. Our ordinary notion of a reason also seems to combine two quite different ideas: what in the lectures I have called the idea of *normativity* or (again there is no non-redundant way to put it) of *obligating* someone to do or believe something or *requiring* someone to do or believe something, and again the idea of universality – that the normativity must be captured in a regular or law-like formulation. What the *normativity* of reasons and the *power* of causes seem to have in common is that they are forms of necessitation: a cause *makes* its effect happen, and so necessitates it (SN 226)

Although, as Korsgaard says, the notions of power and universality may be different, there is an easy conceptual argument to be made connecting them. For some fact, *A*, to cause or determine some other one, *B*, is for *A* being true to rule out all alternatives to *B*. Making something the case means ruling out all possibilities in which it is not the case. But to say that *A* rules out not-*B* is just to say there is a law: if *A*, then *B*. All that law says is *A* and not-*B* cannot be combined, as is true if *A* determines that *B*.

This easy conceptual connection is what allows Kant to say in the *Groundwork*, in no more words than these, that "the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited" (G 4:446). Trivially, all determination involves or is describable by some kind of law. The reason is that what it means to determine something is to rule out all the alternatives.

4.2.2

Kant and Korsgaard both talk of causing – Kant's early modern German for "causality" is "Causalität."[40] I have been talking of determining, and I do have a reason for this. The reason is we now sometimes talk about indeterministically causing things. Where causation and determination come apart, however, Kant and Korsgaard mean determination.

What is meant by saying that one event or fact causes another indeterministically is usually that it determines *that* it happens, but not when or some details of the way. The well-worn examples are of nuclear phenomena. On some physical theories, it would be the case that nothing determines precisely when unstable, radioactive atoms decay. The makeup of that atom before it decays determines that it will decay, and also determines what will go on in the moment when it does decay. Because the when is left open, we might say that the event of the

[40] The modern spelling would be "Kausalität."

decay is determined incompletely. The prior state of the world only partially determines the details or the facts of that event.

Besides the terminological point, a possibility which this example helps bring out is that we might determine only aspects or properties of events. Were we to do this, we would rule out some ways an event could go without ruling out others. You might determine that your club hits the golf ball, but that it strikes it on the sweet spot be largely down to luck. You might determine that you chase your natural desires, but your nature, not your choice, determine what those are.

4.2.3

The fact we have to try to act allows us to assume, just for that purpose, that we have some ways of doing so. If we saw no options whatsoever, then we could not try to act, and yet we must do so. Earlier, I called what we are warranted to assume our "practical conditions" (2.2.4). I also explained the sense in which they have "reality:" they are granted in the scope or for the purpose of trying to act (2.4).

What our having to try to act does not tell us is how fully or completely we can determine our actions. That is, it does not tell us how many alternatives to what we do we can assume we get to rule out through our choices. What we can assume are two claims. The first is that there is *some* gap for our choices to bridge. The state of the world without our choices is not sufficient to completely determine how we act. We can assume that because we only get to choose our actions if there is some choice left for us to make. If we have to try to do some determining, there must be something left for us to determine.

The second claim we can assume is, more obviously, that we are able to bridge whatever gap is left for us. From this it follows that if we *cannot* bridge some size of gap, then it cannot be

ours, or cannot possibly be that big. There might be kinds of fact or event which it is logically or conceptually unintelligible that we could determine. It would follow that what we get to determine is something other than of that kind.

We can see these two claims as two pressures which together might generate some philosophy of action. Action is a porridge – Kant's "delicacy for the savages" (AL 25:685) – which, accordingly, must turn out neither too hot nor too cold. What we determine in choosing our actions must not turn out to be nothing, but it also must not turn out to be so demanding that we could not see ourselves as determining it.

I am about to argue that these two claims can be seen to generate a Kantian constitutivist philosophy of action. The necessary assumption that we can act – both in the sense that there is room for us to, and that we can occupy that room – can be seen to necessitate a particular way of conceiving of what we would do in acting. It is worth repeating firstly, however, that all of this is at the level of practical necessary assumptions. The philosophy of action for which I will be arguing will be a scoped, nonempirical, nonspeculative, *a priori* one.[41]

4.3 Content and incentives

4.3.1

That clarification in hand, we can continue with the argument from the two pressures on a practically adequate philosophy of action. I claimed it is not given that in choosing our actions, we rule out all alternatives to them ourselves. What is given is that there is a space for us to determine something, but not so big we cannot fill it. The way in which a space might be

[41] I do not have space to defend this claim here, but I did so in my 'Kantian indifference about moral reason.'

too big, I suggested, was if *you*, the agent, could not intelligibly fill it. What you determine in choosing your actions must be something which *you* determine through that choice.

To be acting at all, our actions need some content. That content must at least amount to an act: that is, a particular happening, or the lack of one (3.2.1). What I mean by "content" is something which narrows down, or fills in parts of, the choices we might make. For that content to amount to an act would be for it to rule out all acts but that one.

There is a puzzle about what contribution we can intelligibly make to the content of our actions. The only obvious way to get any is by handing the determination of it to something other than our choice. Suppose, for example, you chose to follow your natural inclinations in your actions. There may not even be a way of seeing how you could choose that: why follow your inclinations, and not something else? Suppose you can, however. What you do in choosing to follow your natural inclinations is leave the determination of the content of your actions to your inclinations. To determine that something is the case is to rule out the alternatives to it being the case (4.2.1). If you follow your inclinations, what rule out any alternatives to the way you act are your inclinations. All that needs to change for your actions to be different is for your inclinations to be different. What explains why you act in the particular way you do is only that your inclinations are not somehow different.

The problem is not unique to your natural inclinations. It arises just the same if your source of content is your nature construed more generally, your circumstances, God, "real" moral truths, or the roll of a d20. If you simply follow a source of content, then the ruling out of alternatives to its suggestions is not done by you; and this is all ignoring the problem that your choice of any source would be a choice for which you needed content too.

4.3.2

We need content for our actions, but cannot seem to generate it ourselves. Our only way of coming to have it is by letting it be determined by something else. To put the argument another way: whatever contrastively explains our particular choices determines what their contents are, since determining and contrastively explaining are both the ruling out of alternatives.

Kant and Kantians sometimes talk about "spontaneous" choices. These are choices which are inherently inexplicable, because they are conceived as being self-caused. Nothing, not even luck, would explain why a spontaneous choice was this choice and not another choice. There would be no such contrastive explanation: it would just be some kind of fact spontaneous choices were as they were.

For this reason, the idea of spontaneity cannot help us here in trying to see how we can come to have content. I have been suggesting that to determine something is to rule out the alternatives. In a spontaneous, inexplicable choice, there would be nothing ruling out the alternatives. That means that its content would be determined by nothing, not that you in somehow making the choice would determine it.

Another way to put this is that if we could make spontaneous choices, then they could not be determinations of content. That is, they could not rule in our acting in some ways and rule out others. The most they might conceivably do is rule in or out our acting at all, and even then, in a way which does not count actively doing nothing as not acting at all.

4.3.3

What is clear from the argument I have been making is that the content for our actions must be in some way proposed to us. We must have content for our actions; we cannot generate it

ourselves; but since we must determine something, it cannot be forced on us. It must be in some way offered, with what we determine closing the gap between these proposals and our actions. To work out what those proposals are like, we need to look to what they must be like if we are to somehow end up acting.

Firstly, these proposals must be of whole actions, rather than just parts of actions. That is, they cannot need assembling in some way before we act on them. If these proposals somehow needed assembling, we would need to choose how to assemble them. We would face a further problem of choice to the one which the proposals were recommending answers. In creating further problems of choice, unassembled proposals would create further needs for content. So, these proposals must be preassembled, recommending whole actions, to meet our needs for content.

Secondly, at least *some* of these proposals must be ones on which we could intelligibly act. Suppose proposals of actions could have some property which meant it was conceptually impossible to act on them. Since we already know what is coming: suppose that property was the property of the action being immoral. We must be able to act, and so it would follow that we must always have at least one morally permissible proposal.

4.3.4

These two characteristics I am describing are those of what Kant and Korsgaard call "incentives." What I have been doing just now is showing how Kantian constitutivist premises generate or motivate that concept. As Korsgaard explains it in *Self-Constitution*:

> In Kantian moral psychology, the starting point for action is what Kant calls an incentive (*Triebfeder*). An incentive is a motivationally loaded representation of an object. I am using the term "object" broadly here to include not only substances but

> also states of affairs and activities. The object may be actually perceived, or conceived as a possible item in the environment, a way that things might be. You are subject to an incentive when you are aware of the features of some object that make the object attractive or appealing to you. (SC 104–05)

Korsgaard can make it sound as if, for her moral argument to get off the ground, we need to accept a Kantian moral psychology, with that being the philosophical starting point. What I am trying to make clear in this chapter is that this does not need to be the starting point. There are ways to step from some very general claims about our trying to act to a rich Kantian moral psychology.

In a later discussion of rational creatures' "liberation from the control of instinct," Korsgaard further explains:

> Instincts still operate within us, in the sense that they are the sources of many of our incentives – in fact, arguably, though by various routes, of all of them. But instincts no longer *determine* how we respond to those incentives, what we do in the face of them. They *propose* responses, but we may or may not act in the way they propose. Self-consciousness opens up a space between the incentive and the response, a space of what I call reflective distance. It is within the space of reflective distance that the question whether our incentives give us reasons arises. (SC 116)

Here incentives are framed as proposals by our instincts, where what Korsgaard means by our instincts are the psychological forces at work in us which make acts and ends seem fitting or appealing.

4·3·5

I said a moment ago that we must always have at least one proposal on which we could intelligibly act. If it does turn out that all action, strictly speaking, must be moral, this would mean that we must always have a moral incentive. In the second *Critique*, Kant makes the claim that:

> if by *incentive* (*elater animi*) is understood the subjective determining ground of the will of a being whose reason does not by its nature necessarily conform with the objective law … the incentive of the human will (and of the will of every created rational being) can never be anything other than the moral law (C2 5:71–72)

Similarly, in the *Religion*, Kant says:

> the moral law is itself an incentive in the judgment of reason, and whoever makes it his maxim is *morally* good. Now, if the law fails nevertheless to determine somebody's free power of choice with respect to an action relating to it, an incentive opposed to it must have influence on the power of choice of the human being in question (Rel 6:24)

We always have a moral incentive. If we somehow fail to act immorally, then we must have had more than just the moral incentive.

5 Why should our moral psychology feature ends?

5.1 Two arguments from spontaneity

5.1.1

I have been arguing that how we see ourselves in having to try to act justifies the concept of incentives. These incentives are proposals of whole actions, offering to meet our needs for content. What needs to be worked out now is how we could determine ourselves to act on any incentives. That is, we need to see how we could determine our responses to our incentives.

A moment ago, I briefly introduced the Kantian notion of a spontaneous choice (4.3.2). I said that such a choice would not be a way for us to determine our actions' content. The position of our having no content we can take for granted is sometimes described as the position of spontaneity. It is that of our having no clear way to rule out any actions, or rather any of our incentives. We have no clear way to do that because distinguishing between our incentives would seem to require us to already have some content.

Korsgaard attempts an argument from this position of spontaneity to Kant's categorical imperative. It turns on the claim I made earlier, that all determination involves a law. In 'Morality as freedom,' she explains:

> Although incentives do not yet provide reasons for the spontaneous will, they do determine what the options are – which things, so to speak, are candidates for reasons. … In the *Religion,* Kant claims that it is impossible for a human being not to be moved at all by incentives; our freedom, rather, is exercised in choosing the order of precedence among the different kinds of incentives to which we are subject. So the

> real choice will be between a maxim of self-love, which subordinates the incentives of morality to those of inclination, and the moral maxim, which subordinates incentives of inclination to moral ones. (CKE 165)

The "moral maxim" here is roughly that of acting on whatever incentives we have unless doing so would violate Kant's Formula of Universal Law. Given the connection between determination and rules or laws, the spontaneous will's search for a way to determine itself can be understood as a search for a law. As Korsgaard understands it, however, all Kant's Formula says is that the spontaneous will must choose a law:

> This formula merely tells us to choose a law. Its only constraint on our choice is that it have the form of a law. Nothing provides any content for that law. *All that it has to be is a law*. (CKE 166)

In following that Formula:

> the free will retains the position of spontaneity. Or, to put it a better way, the argument shows that the free will need do nothing to make the Formula of Universal Law its principle: it is already its principle. … The will that makes the categorical imperative its law merely reaffirms its independence of everything except law in general. Its dependence on law in general is not a constraint, for that is just a consequence of the fact that it is a will. (CKE 166–67)

In other words, Korsgaard believes we find in Kant an argument that we could not intelligibly determine ourselves to violate the categorical imperative. All the Formula of Universal Law represents, she believes, is the kind of determination which would be possible for us from a position of spontaneity. That being the position which we find must assume ourselves to be

in, we must assume that in trying to determine our actions, we would be making a mistake in trying to violate Kant's Formula.

5.1.2

Korsgaard's argument in this paper is quick. Firstly, as I have noted and as in other places, Korsgaard does not give an argument for why we ought to conceptualise our moral psychology in terms of incentives. That was the point of the previous chapter: to fill in that missing argument. The concern I want to press about Korsgaard's argument now is something different.

That concern is that Korsgaard moves quickly from determination necessarily involving *some kind* of rules or laws to the only way we can intelligibly determine anything being according to *universal* laws. It is a common worry with these sorts of Kantian arguments that no real explanation is given for why I cannot, say, just make laws for me. But Korsgaard is just being quick, the same way Kant is quick when he says "the concept of causality brings with it that of laws" (G 4:446). As with Kant's claim, there is a simple argument from the one idea to the other.

What lets Korsgaard – and Kant – move from rules or conditionals to universal laws is that the choice being imagined is being taken from a position of spontaneity. I described that position as that of having no content we can take for granted and make our choices with. Our actions need content, but to get any content for them, we need to choose between possible sources of content. We already need some kind of content – some basis – however, to choose between those possible sources. We cannot just make that choice magically, spontaneously, because without some explanation why we did not choose differently, that choice cannot intelligibly be seen as a ruling-out: a determining.

That was the ground covered so far. There is a problem seeing how, from a position of spontaneity, we can take our first step towards choosing our actions, by narrowing down their content. Another way to think of the problem is that, from this position, we have nothing particular to go on. That includes anything particular about ourselves, our interests, and our circumstances.

In at first a trivial sense, however, this means the position of spontaneity is one of universality. Choices made from this position can only be universal because they have no way of being particular. As Kant puts it in the second *Critique*, "a rule is objectively and universally valid only" – and exactly – "when it holds without the contingent, subjective conditions that distinguish one rational being from another" (C2 5:21). Granted, it is hard to see how any contentful choice, particular or otherwise, could be made from the position of spontaneity. This is not a question Korsgaard takes up in 'Morality as freedom:' her argument ends at universality. What her argument shows, if it works, is that *if* a will can somehow choose to act on some of its incentives, then those incentives must be the ones which cohere with Kant's Formula of Universal Law. They must do so because what that Formula reflects or represents is that a will's first or highest-level choice, before it has any content, cannot be particular.

5.1.3

Of course, Kant's argument in the second section of the *Groundwork* does not end with the Formula of Universal Law. Similarly, his argument in the second *Critique* does not end with the "Fundamental Law of Pure Practical Reason" (C2 5:30). In the *Groundwork*, Kant moves on to the Formula of Humanity and then the Formula of Autonomy. In the second *Critique*, he also goes on to develop the notion of the Highest Good. To show how the concept of an end can be motivated or generated in a Kantian constitutivist philosophy of action, I will need

to make an argument from the position of spontaneity to something like Kant's "kingdom of ends" (G 4:433).

Such an argument might begin this way. In the position of spontaneity, there are two key points we know. The first is that we must act as proposed by some incentive. The second is that our choice to do so cannot be based on any content, and so on anything particular. It might seem as if our own incentives have some kind of privileged status as the ones which we must ultimately act upon. They are by themselves sufficient to meet our needs for content: also looking to others' incentives seems unwarranted. All that our need to act can surely justify is that we somehow draw on our own incentives. As intuitive as that might seem, however, we can distinguish between two questions.

The first is that of where the content for your actions must immediately or proximally come from. We know the answer is your incentives: your incentives were functionally defined as what answer that question. Wherever that content comes from ultimately, it must come to you via your own incentives. And this is the second question: where that content ultimately comes from. Earlier I argued for two characteristics of incentives, the second of which was that they stretch, in a sense, to meet our needs (4.3.3). If you have some set of incentives but cannot intelligibly act on any of them, that set cannot be complete. You must have more incentives than those: you must have at least one you can intelligibly act upon. If there was some reason why you needed to take account of more than just your own incentives, then your incentives would adapt to meet that need.

In other words, the question of what incentives we must ultimately see as sources of content for ourselves is not answered by the fact that we must end up acting on our own incentives. There are three overlapping pools of content here. The first and smallest is that of your own

ordinary incentives: the ones which you have naturally, so to speak, and are not generated by your need to have incentives which meet certain conditions. The second, of middle size, is that smallest pool plus any incentives you must have because – it was supposed (4.3.5) – you need to act morally, and would not otherwise have any moral incentives. The third, potentially much larger pool is that of whatever content you must ultimately take account of in choosing your actions. The content of this third pool would determine the difference between the first and second.

What I was arguing a moment ago is that your need to act on your own incentives entails that the content of your actions must come from incentives in the second pool. It does not entail that the third pool must be no bigger than the first: that is, that the only incentives which matter are those which you yourself have naturally. In fact, it may not seem as if your own natural incentives need to be in the third pool at all. If that third pool can generate its own incentives in the second, it can by itself provide you with content to act on.

5.1.4

The third pool does include the content of the first, and we can see why in the following way. As we have seen already, there is no obvious general way for anything to earn the right to be a source of content for you. That means that the third pool is empty by default. This is the problem in the position of spontaneity: you have no obvious avenue to determining yourself to take account of anything. We know that you must act, however, and we know that it must be in ways proposed to you by your incentives. Provisionally, then, the third pool must include your own incentives. In the absence of a way for the third pool to generate its own incentives, there is nothing left to act on but your own natural incentives.

However, this inclusion of your own natural incentives in the pool of content of which you must ultimately take account must be done in a way which is intelligibly possible for you from the position of spontaneity. It is you who have to choose to take account of anything, and so you must be able to see yourself as doing so before you have any intelligible basis for taking particular interest in your own ends and situation. Provisionally, you have a need for your natural incentives to be in the third of the pools, but it does not follow that your way of including them can be particular to you. The nonparticular, or universal way to take account of at least your own natural incentives would be to take account of all incentives: not just your own, but other beings' too.

Suppose that your own natural incentives, the first pool, could be united or made coherent in some way. Kant gives such a union the label "Glückseligkeit," usually translated as "happiness." In the second *Critique*, Kant considers what could be the matter or the content of a moral will. He supposes a starting point:

> Let the matter be, for example, my own happiness. This, if I attribute it to each (as, in the case of finite beings, I may in fact do), can become an *objective* practical law only if I include in it the happiness of others. Thus the law to promote the happiness of others arises not from the presupposition that this is an object of everyone's choice but merely from this: that the form of universality, which reason requires as the condition of giving to a maxim of self-love the objective validity of a law, becomes the determining ground of the will; … only from this limitation, and not from the addition of an external incentive, could there arise the concept of obligation to extend the maxim of my self-love to the happiness of others as well. (C2 5:34–35)

Kant's thought here can be unpacked in the following way. In the position of spontaneity, what you determine must be something you could see any rational being as determining. That follows from the fact that you cannot distinguish your situation from that of other rational beings. You having nothing particular to go on which could make your determinations different to anyone else's. What Kant effectively says in the last sentence above is that your third pool starts off as empty. That means it only has whatever content you determine to be in it. That determining by you, however, must be the same for any rational being. It follows that what I have called your third pools, your sum of content to take account of, must be the same.

5.2 Reconciling incentives

5.2.1

Now, if anything, this should make it *more* difficult to see how we can work our way down to particular actions. This will be part of the point: it will be part of what we need ends for. We are yet to see what we ourselves could intelligibly determine about our actions, and now there seem to be impossible demands on their content. Somehow, it needs to be possible to combine the collected incentives of all sentient beings, according to Korsgaard, or of all rational ones according to Kant.

Korsgaard differs from Kant here because she argues that the fact our third pools must be the same as other rational beings' does not actually justify, or make intelligible, restricting our concern to the incentives of only rational beings. That is, the fact that it is rational beings with whom we must share the choice that gets us any content does not entail that the content picked up by that choice must come only from rational beings. Korsgaard believes that all animals can be understood as being subject to incentives, and so there is a question why a nonparticular choice would not range over their incentives.

In her *Fellow Creatures*, Korsgaard puts the point in terms of ends-in-themselves. Roughly, a being which was an end-in-itself would be one whose incentives we needed to consider. To presuppose your own value, in the quote that follows, is to take yourself to be an end-in-yourself. Korsgaard considers why Kant thought only the incentives of rational beings matter:

> One possible reason, I suppose, is that it is only rational beings who must presuppose our own value in order to engage in practical activity. This, however, would be an inadequate reason. The idea that rational choice involves a presupposition that *we* are ends in ourselves is not the same as the idea that rational choice involves a presupposition that *rational beings* are ends in themselves, for we are not merely rational beings. We are also animals – beings who have a good. Of course, the other animals don't have to presuppose anything in order to engage in practical activity, since they are not rational beings who need to be able to endorse their choices or to see them as justified. But the *content* of the presupposition behind rational choice is not automatically given by the fact that it is only rational beings who have to make it. (FC 142)

You have a need, provisionally, to include your own natural incentives among your sources of content. In the position of spontaneity, however, you have no grounds for stopping at your own incentives. What Korsgaard argues, contra Kant, is that just as you lack grounds for stopping at your own incentives, you also lack grounds for stopping at just rational beings' incentives.

Whoever's argument we follow, we still find ourselves with an unwieldy mass of pooled incentives. Where we once had no content, the assumption we can act on our incentives, universalised, now means that we have too much content. This is where ends come in: not just

to the moral argument, but to the Kantian constitutivist moral psychology. We must suppose that incentives are in some way reconcilable, and therefore we must suppose they include not just acts, but ends.

5.2.2

In Book II of the *Eudemian Ethics*, we find Aristotle discussion what it means to make a decision (*prohairesis*), or what kind of object a decision is. He concludes that a decision is "clearly … a deliberative desire for things that are up to oneself," where a deliberative desire is "one whose starting point and cause is deliberation" (EE 1226b). On the view he sketches:

> The soul's deliberative power is its capacity to apprehend some cause, since that for the sake of which is one of the causes. Cause is a reason why, and that for the sake of which a thing is or comes about is, we say, a cause. For example, collecting one's money is a cause of one's walking, if one is walking for the sake of that. That is why those who have no aim before them are not liable to deliberate. (EE 1226b)

Causes and aims, in this passage, might be thought of as both the grounds of actions, but seen in opposite directions. To apprehend the cause of something is to first see it, and then to see behind it its aim or ground. When you have an aim and are deliberating, what you first see is the ground on which you will be acting. What you then deliberate to, forwards, are the actions which could have that cause or ground.

The argument of the passage therefore runs like this. Having the power to deliberate involves having the capacity to see the grounds or the aims of doing things. You have to be able to consider what aims would be served by particular actions. That new information, those aims, can be the basis of intelligent distinctions between those possibilities. For those distinctions to lead anywhere, however, you need to have some aims before you. If you are indifferent

between the aims which might be served by your options, being able to apprehend them does not get you anywhere. To be able to deliberate, you need both to have some aim, and to be able to apprehend what would fit it. Only then can you narrow down the actions you might take.

5.2.3

What we need to do in the position of spontaneity is narrow down the proposals made my incentives. As Aristotle's argument suggests, what we need to do this are aims, or, in more Kantian terms, ends. Moreover, there are two respects in which we need these aims or ends. Firstly, we need our incentives – our proposals of actions – to come with ends. This would allow us to distinguish between incentives on the basis of ends. Secondly, we need to have some ends "before us," in the sense that we must have some ends on the basis of which we can rule out contrary, or merely uninteresting, incentives.

Hopefully this seems like familiar territory. Our need for ends and their being included in incentives will mean that ends, as well as acts, become collected in our pools of incentive-proposed content. Like the acts originally, the ends in that pool may then conflict: we may need to choose between them. There will be no obvious way out but to go up a level again, to ends of ends. This is how the argument must go, and on through ends of ends of ends. I will not try to spell out these cycles, since their number is not guaranteed. There will be a resolution, however, because ends, unlike acts, do not have to conflict. That is, one act, or a range of acts, can be supported by multiple ends. What we are doing in going up these levels, or through these cycles, is trying to be general enough about some ends that there end up being some acts they can be seen to support together.

The easiest way to grasp the mechanics of this process may be to imagine it with just two people. Imagine there are, for the purposes of this Kantian constitutivist moral argument, just two people with incentives. To simplify things further still, suppose they have only a single natural, "first pool" incentive each. The implications of their having more incentives would be the same for this purpose as there being more people.

The first person's incentive says: play my music loudly in Garden Quad, for the sake – the end – of enjoying my music where I am, in Garden Quad. The second person's incentive says: work hard in my room in the staircase next to Garden Quad, for the sake of being well-prepared for my exam the following morning. The acts are the playing the music and the working hard, in their particular locations. I have given what might be thought of as the first-level ends for those acts: enjoying your music where you are, and being well-prepared.

The acts themselves are obviously different, and that is all there is to be said at that level. Two people are proposing to manipulate different bodies – their own – to do different things in different locations. At the level of ends, there is also a difference, and as I have spelled things out, one we can suppose is irreconcilable. The second person cannot work hard with the music of the first playing loudly out their window. There is a genuine conflict of ends at this level of particularity or specificity. So, we must go up another level, and ask what ends sit behind or underneath those which I have given these two people as pursuing.

The first person wants to enjoy their music where they are because they want to enjoy their music, and they want to do that because they want to celebrate and destress after their own exams. We can think of there as being a chain here: from enjoying their music in Garden Quad, back to enjoying music, celebrating and destressing, and behind that what we might just call having fun. The person wants to have fun because, to put things a little redundantly,

they enjoy it, and they value their enjoyment of things because, ultimately, they place some kind of value on themselves.

Or so someone like Korsgaard will argue. The second person wants to work hard because they want to be well-prepared for their exam, for the sake of doing well in it, so as to have strong grades when they leave university. The pair's ends stop conflicting as soon as it is no longer given that one or the other needs to be in the area of Garden Quad. It never needs to be called into question whether the best way for the first person to celebrate and destress is with music, since, we will suppose, places can be found for them to do that still conveniently. It is not given from what I have said that the person who needs to move is the first, not the second, but there is now scope to work down from these higher-level, reconcilable ends to acts again.

5.3 The Kantian constitutivist picture

5.3.1

The purpose of this chapter was primarily to motivate the inclusion of ends in the Kantian constitutivist philosophy of action. It was not to defend the moral arguments which I have needed to sketch along the way. Those arguments were needed because I think what justifies the inclusion in that philosophy of action of ends is that ends are necessary for our incentives to be reconcilable: for there to be a prospect of a kingdom of ends.

Kant's hope is that if we take all the chains of ends leading up from our incentives, there will be conceivable "a whole both of rational beings as ends in themselves and of the ends of his own that each may set for himself" (G 4:433). If we take all those chains, knock out those links which necessarily conflict, and work back down again, we will be able to form something imaginable as a common, shared tree of ends. The lowest, thinnest branches will be our projects at their most particular, and our particular ways of pursuing them will stand to be

respected if they do not conflict with others'. The thickest bough, or the trunk, will be every being valuing itself for its own sake, where what that means is "that we each take the things that are good-for us to be good absolutely" (FC 148).

5.3.2

A question it would be too much to leave is that of what action ends up being on this Kantian constitutivist picture. What is it that we end up determining in our actions, and how is our doing so intelligible? The argument I have been giving would give us content for our actions. We can pursue our own incentives unless they do not fit into the tree, or the kingdom, I have just been describing. In that case, that tree of ends would generate a further incentive we could pursue. This may sound odd, but remember that on this picture, actively doing nothing is a kind of acting (3.2.1). Morality can tell us to just *keep thinking* until we find something moral to do.

What it might seem is left for us to determine, then, is which of our incentives to pursue when there remain multiple which we can permissibly. Strictly, however, this is not the case. Something needs to whittle those incentives down to one. This will be the work of what is sometimes called "nonmoral" or "prudential" practical reasoning: you deciding what you would *most like* to do out of the permissible options you have. An interesting point about the kind of Kantian argument I have been describing is that it gives no reason to think of this prudential reasoning as a kind of determining. In Kantian terms, the freedom which we have from the practical standpoint is not a kind of latitude. When you are whittling down your options, trying to leave yourself with one which appeals to you above the others, although this may *feel* like some kind of free process, the argument gives no basis for thinking of it that way.

Once the options which conflict with the common tree of ends have been ruled out, and prudential reasoning has left you with one live incentive, all there is left to do is to determine yourself to act on that incentive. This will not be a choice between options, to determine or not determine, since the incentives are the options. Equivalently, you cannot determine yourself not to determine anything. Not being a choice between options, moreover, there are no alternatives left to be ruled out by content. There is therefore no problem seeing how you can determine yourself to take the action rather than anything else.

What you determine in your actions, then, turns out simply to be that you act at all. This is all you can intelligibly contribute, once all the contents of your action have been determined by their sources through your incentives. What we can intelligibly determine is that we assent when we have one option left. The "free" part of your action, in the Kantian sense, turns out to be a buttonpress.

5.3.3

Kant and Korsgaard both have stories about how we can be held responsible for immorality even though, from our own practical perspectives, our determining ourselves to be immoral is unintelligible. For Kant at least, part of that story is that something being unintelligible from that perspective does not entail that it could not have happened and have been in some sense our doing. Suppose we can imagine a moment in which we lacked all the conceptual framework which gets these moral arguments going. If we *did* something in that moment, could it conceivably have set us up to fall into immorality when we fail to determine our actions now, with that framework?

As I read him, Kant argues something like this in the *Religion*. His argument has its obscurities, and brings in ideas which I have not here.[42] The reason why I bring it up is to emphasise the status of the claims I have been making in these chapters.

Claims about what we can intelligibly determine here are ultimately about what we can see as our options. We see ourselves as faced with the problem of having to try to determine how or that things happen. It turns out that as we conceive of that problem, the only things we can determine ourselves to do are moral things. As such, in the scope of that trying to determine, we ought to do the moral things.

Doing so, in other words, is how to do what it is we are trying to do. The "ought" carries the force of our own, inescapable projects as rational beings. We are trying to be agents, and what the constitutivist – any kind of constitutivist – tries to do is to show us what to do in order to succeed.

[42] I considered the argument in my 'Kant's account of human dignity.'

Bibliography

I have used inline references for Korsgaard's works and historical texts. In alphabetical order by the abbreviations used for them, those sources are:

AL: Kant, Immanuel. 2012. *Lectures on Anthropology*. Edited by Allen W Wood and Robert B Louden. Translated by Robert R Clewis, Robert B Louden, G Felicitas Munzel and Allen W Wood. Cambridge: Cambridge University Press.

Anth: Kant, Immanuel. 2006. *Anthropology from a Pragmatic Point of View*. Translated by Robert B Louden. Cambridge: Cambridge University Press.

AR: Korsgaard, Christine M. 2009. "The Activity of Reason." *Proceedings and Addresses of the APA* 83 2: 23–43.

AT: Descartes, René. 2013. *Meditations on First Philosophy*. Translated by John Cottingham. Cambridge: Cambridge University Press.

C1: Kant, Immanuel. 1998. *Critique of Pure Reason.* Translated by Paul Guyer and Allen W Wood. Cambridge: Cambridge University Press.

C2: Kant, Immanuel. 1996. *Practical Philosophy*. Translated by Mary J Gregor. Cambridge: Cambridge University Press.

CKE: Korsgaard, Christine M. 1996. *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press.

COA: Korsgaard, Christine M. 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

Conj: Kant, Immanuel. 2006. "Conjectural Beginning of Human History." In *Toward Perpetual Peace and Other Writings on Politics, Peace, and History*, edited by Pauline Kleingeld, translated by David L Colclasure, 24–36. London: Yale University Press.

EE: Aristotle. 2012. *Eudemian Ethics.* Translated by Brad Inwood and Raphael Woolf. Cambridge: Cambridge University Press.

FC: Korsgaard, Christine M. 2018. *Fellow Creatures: Our Obligations to the Other Animals.* Oxford: Oxford University Press.

G: *Same as C2*.

LEP: Diogenes Laertius. 1972. *Lives of Eminent Philosophers.* Translated by R D Hicks. Cambridge, MA: Harvard University Press.

M: *Same as C2*.

MMVS: Korsgaard, Christine M. 1998. "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind." *Ethics* 109 1: 49–66.

NG: Korsgaard, Christine M. 2011. "Natural Goodness, Rightness, and the Intersubjectivity of Reason: Reply to Arroyo, Cummiskey, Moland, and Bird-Pollan." *Metaphilosophy* 42 4: 381–94.

Pro: Kant, Immanuel. 2004. *Prolegomena to Any Future Metaphysics.* Translated by Gary Hatfield. Cambridge: Cambridge University Press.

Rel: Kant, Immanuel. 1996. *Religion and Rational Theology.* Translated by Allen W Wood and George di Giovanni. Cambridge: Cambridge University Press.

REM: Korsgaard, Christine M. 2010. "Reflections on the Evolution of Morality." *The Amherst Lecture in Philosophy* 5: 1–29.

Rep: Plato. 2012. *Republic.* Translated by Christopher Rowe. London: Penguin.

RK: Korsgaard, Christine M. 1995. "Rawls and Kant: On the Primacy of the Practical." Edited by Hoke Robinson. *Proceedings of the Eighth International Kant Congress.* Milwaukee: Marquette University Press. 1165–73.

SC: Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity.* Oxford: Oxford University Press.

SN: Korsgaard, Christine M. 1996. *The Sources of Normativity.* Cambridge: Cambridge University Press.

The other works cited are:

Allison, Henry E. 2004. *Kant's Transcendental Idealism.* London: Yale University Press.

Campbell, C A. 1951. "Is 'Freewill' a Pseudo-Problem?" *Mind* 60 240: 441–65.

Darwall, Stephen. 2000. "Normativity and Projection in Hobbes' Leviathan." *The Philosophical Review* 313–47.

Enoch, David. 2006. "Agency, shmagency: why normativity won't come from what is constitutive of action." *The Philosophical Review* 115 2: 169–98.

Enoch, David. 2009. "Can there be a global, interesting, coherent constructivism about practical reason?" *Philosophical Explorations* 12 3: 319–39.

Enoch, David. 2011. "Shmagency Revisited." In *New Waves in Metaethics*, edited by Michael Brady, 208–33. Basingstoke: Palgrave Macmillan.

Ferrero, Luca. 2009. "Constitutivism and the Inescapability of Agency." In *Oxford Studies in Metaethics*, vol. IV, edited by Russ Shafer-Landau, 303–33. Oxford: Oxford University Press.

Gibbard, Allan. 1999. "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics* 110 1: 140–64.

Gibbard, Allan. 2003. *Thinking How To Live*. Cambridge, Massachusetts: Harvard University Press.

Hill, Thomas E Jr. 1992. *Dignity and Practical Reason in Kant's Moral Theory*. London: Cornell University Press.

Hussain, Nadeem J Z, and Nishi Shah. 2006. "Misunderstanding Metaethics: Korsgaard's Rejection of Realism." In *Oxford Studies in Metaethics*, vol. 1, edited by Russ Shafer-Landau, 265–94. Oxford: Oxford University Press.

Hussain, Nadeem J Z, and Nishi Shah. 2013. "Meta-ethics and its discontents: a case study of Korsgaard." In *Constructivism in Ethics*, edited by Carla Bagnoli, 82–107. Cambridge: Cambridge University Press.

Katsafanas, Paul. 2013. *Agency and the Foundations of Ethics: Nietzschean Constitutivism.* Oxford: Oxford University Press.

Lehrer, Keith. 1960. "Can We Know That We Have Free Will by Introspection?" *The Journal of Philosophy* 57 5: 145–57.

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2004. "The Phenomenology of Free Will." *Journal of Consciousness Studies* 11 7–8: 162–79.

Rawls, John. 1980. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77 9: 515–72.

Roberts, Adam J. Essay for assessment. "In defence of Korsgaard's arguments for human and animal value." King's College London.

Roberts, Adam J. Dissertation for assessment. "Why be a Kantian agent." King's College London.

Roberts, Adam J. Essay for assessment. "Kant's account of human dignity." University of Oxford.

Roberts, Adam J. Essay for assessment. "The missing two-thirds of metaethics." University of Oxford.

Roberts, Adam J. Forthcoming/Essay for assessment. "Kantian indifference about moral reason." *Behavioral and Brain Sciences*/University of Oxford.

Searle, John. 1984. *Minds, Brains, and Science.* Cambridge, MA: Harvard University Press.

Street, Sharon. 2010. "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 5: 363–84.

Velleman, J David. 2009. *How We Get Along.* Cambridge: Cambridge University Press.