

Make Up Your Mind: Reflection and the Necessity of Agency

Examination and Defense of Korsgaard's Constitutivism

'What is it that will result from this right action of ours?'
Immanuel Kant

Master's Thesis for the
Research Master Philosophy
Utrecht University
June 20th, 2020

By: Bram Galenkamp

Student number: 5502322

Word count: 39,612 (excluding abstract, acknowledgements and table of contents)

Supervisor: Dr. Joel Anderson

Second reader: Prof. Dr. Micha Werner

Third reader: Dr. Niels van Miltenburg



Universiteit Utrecht

For Maartje, whose inspiration and support have made all this possible.

Contents

Abstract	4
List of abbreviations for cited works	5
Introduction.....	6
Chapter 1: Korsgaard's constitutivism	8
1.1 Action, agency and self-constitution.....	9
Reflective endorsement	10
Reflective endorsement narrowed down: contradictions in the will.....	14
Agency and its constitutive principles.....	17
Agency revisited	21
1.2 Constitutivism and normativity	23
Chapter 2: The problem with agency	27
2.1 Enoch's shmagency objection	27
2.2 A Korsgaardian reply to Enoch	30
Morality as a vocation	31
2.3 Anthropological necessity	38
The hypothetical status of desires	43
The importance of wholeheartedness	46
The underdevelopment of shmagency	51
2.4 The indirectness-objection.....	53
The shmagency trilemma	55
2.5 Anthropological necessity and the shmagency trilemma	57
Skewed and unfair debates	60
Chapter 3: Necessity and normativity	63
3.1 Overstating the importance of universal validity.....	63
Extending hypothetical imperatives.....	64
The universality of reasons.....	65
3.2 The Categorical Imperative and estrangement.....	66
3.3 Autonomy and identity: moral saints.....	69
Conclusion	73
Acknowledgements	76
Bibliography.....	77

Abstract

Christine Korsgaard's recent work focuses mainly on constitutivism about moral normativity and personal identity. In fact, she tries to show that those two things are subject to the same principle, which she identifies as Kant's categorical imperative. Roughly put, the idea is that being a good person is the same as being good at being a person. The latter means that one generally succeeds at finding the right order among one's desires and motives and this, Korsgaard seems to think, is possible only by adhering to the (i.e., Kant's) moral law. Korsgaard's argument is, of course, controversial. In this thesis, however, I will attempt to explain and defend rather than criticize it. One of the main reasons for this is that Korsgaard has not written many responses to criticisms, even though they at times raise real questions about her work. One such question, I believe and will try to show, is posed by David Enoch and is commonly known as the 'shmagency objection'. After providing an in-depth interpretation of the different key elements and their connection in Korsgaard's work, I will address Enoch's objection. I will explore what arguments Korsgaardian constitutivism has at its disposal to counter Enoch's critique, trying to explain why its best hope of showing the necessity of agency lies in what I will call 'anthropological necessity'. That is, I will try to show that, according to Korsgaard, humans are reflective beings and that this analysis of human nature may help show why agency is necessary. In the last chapter, I will explore some potential criticisms and weaknesses of my Korsgaardian approach, such as a demandingness objection and a worry about alienation.

List of abbreviations for cited works

Where possible, the Chicago (17th edition) full note style will be used. For a number of frequently cited works, I will only list an abbreviated title and page number, and an author where necessary. Complete biographical information can be found in the bibliography.

Works by Christine Korsgaard:

- CA *The Constitution of Agency*
- CV “Constitutivism and the virtues”
- KE *Creating the Kingdom of Ends*
- SC *Self-Constitution*
- SN *The Sources of Normativity*

Works by Immanuel Kant:

- C1 *Critique of Pure Reason*
- C2 *Critique of Practical Reason*
- G *Groundwork for the Metaphysics of Morals*
- MM *The Metaphysics of Morals*
- R *Religion within the Boundaries of Mere Reason*

Introduction

Why be moral? It is a question as old as philosophy itself, and judging from the wide array of answers, it is not an easy one to answer. One philosopher brave enough to try her hand at it is Christine Korsgaard, whose work will be the topic of this thesis. In a somewhat distinctive yet rich blend of meta-ethics and normative theory she argues that morality is normative for us human beings because its principles arise from our own will, and that these principles enable us to do what we are ‘condemned’ to do: engage in reflection and action.¹ Morality, according to Korsgaard and fellow constitutivists, is in fundamental respects a lot like non-moral activities, such as building a house. Constitutivism starts with an idea that is as simple as it is ambitious: there are certain standards that establish what makes a good house and what makes a house a house to begin with.² These standards are *constitutive* of houses. Insofar as a building meets these standards, it may be considered a house, and once you are interested in building a house, the constitutive standards are normative. And, constitutivism holds, something quite similar goes for morality: if we want to understand what good actions are, we need to better understand what actions are. Knowing the constitutive standards of action or agency will tell us what action is, what actions are good and thus it will tell us what to do morally.

The merits of this constitutivist approach to morality, I hope, will become apparent over the course of this thesis. One might, however, wonder why an in-depth discussion of Korsgaard’s theory in particular is necessary. She has written a number of books and papers and claims that she essentially always argues for the same thing,³ so shouldn’t her point be clear enough? I believe not. Korsgaard’s writings tend to be quite accessible in terms of style, but her theory contains a number of different elements that tend to stand in a complex relation to one another. Neither the building blocks nor their relation is always clearly presented and perhaps this explains why her theory, as I will argue later, is often misunderstood. Furthermore, Korsgaard seems to have built a bit of a reputation of not responding to critique.⁴ Enoch in particular has a point here, because his “shmagency” objection has stirred up quite a debate, and even though the objection seems to pose serious questions over Korsgaard’s theory, Korsgaard has not written an explicit response. Perhaps these two issues are connected, but rather than speculating on whatever may be lying underneath here, suffice it to say that they invite the following research question: How can Korsgaard’s constitutivism counter Enoch’s shmagency objection?

¹ See, for example, SN chapter 3.

² ‘Constitutivism’ here and at later points in the text refers to Korsgaard’s constitutivism unless otherwise indicated. Other influential constitutivist theories are put forth by David Velleman and Connie Rosati, for instance (respectively in *How We Get Along* (Cambridge; New York: Cambridge University Press, 2009) and “Agency and the Open Question Argument,” *Ethics* 113, no. 3 (April 1, 2003): 490-527).

³ SC 76.

⁴ David Enoch, “Shmagency Revisited,” in *New Waves in Metaethics*, ed. Michael Brady (New York: Palgrave Macmillan, 2011), 208-233, 218; William J. FitzPatrick, “How Not to Be an Ethical Constructivist: A Critique of Korsgaard’s Neo-Kantian Constitutivism,” in *Constructivism in Ethics*, ed. Carla Bagnoli (Cambridge: Cambridge University Press, 2013), 41-62, 45f.

The research question can be split up into three sub-questions, the first being a preliminary question, the second pertaining to Enoch's objection and the third inquiring into some of the potential problems of the answer to the second:

1. What elements comprise Korsgaard's theory and what is their connection?
2. How can those elements be utilized to refute Enoch's objection?
3. What are the potential normative costs of this way of answering Enoch's critique?

One further preliminary remark is in place here. Enoch's objection dates back roughly fifteen years now, and I cannot substantiate this, but the topic does not seem to be raising as much attention as it used to. As should become clear in chapter 2 in particular, this is constitutivism's loss, because the shmagency objection is a serious and fundamental objection. Other constitutivists have tried to answer it, but even though it is much needed, there has not been a distinctively Korsgaardian constitutivist answer.

My attempt to take up this glove will be structured as follows. In the first chapter I will present several key elements of Korsgaard's theory, focusing especially on reflective endorsement, agency and the principles Korsgaard claims are constitutive of agency: autonomy and efficacy. Chapter 2 will involve a discussion of Enoch's shmagency objection. I will first outline what the problem is and then discuss one paper by Korsgaard as a potential response. Then I will argue why it fails and how the elements I have discussed in chapter 1 may be put together to provide a different yet Korsgaardian answer to Enoch's objection, which places the necessity of agency in what I will call 'anthropological necessity'. To conclude the second chapter, I will discuss two different possible interpretations of the shmagency objection and argue that they will not turn the objection into one that can satisfactorily deal with anthropological necessity. In the last chapter, I will briefly go into some of the problems and potential normative costs attached to my Korsgaardian response to Enoch. I will end chapters 2 and 3, respectively, with a brief discussion of the shmagency debate in relation to Korsgaard's theory and the debate around Korsgaard's constitutivism in general.

Chapter 1: Korsgaard's constitutivism

As a starting point, Korsgaard seeks to establish the sources of normativity in the constitutive standards of agency,⁵ although much of what she says also applies to, or cannot do without, the concept of action. Yet in spite of this crucial role for action, agency is the ultimate object of Korsgaard's research. As will be explained and discussed in due course, Korsgaard thinks that we, qua beings with rational capacities, are faced with the task of making something of ourselves.⁶ The project of her recent work then is to show that this 'making something of ourselves' requires principles and that these principles of self-constitution are in fact identical with the most basic principles of morality, allowing her to conclude that, in the end, the moral law is identical to the principle with which we constitute ourselves.⁷

In order to find an answer to my first sub-question and thus to explore Korsgaard's theory and pave the way for a defense against several strong criticisms, I will divide this chapter into two parts: the first focusing on agency and its constitutive principles, the second on the normative side of the theory. Many other divisions are possible,⁸ and mine is of course an artificial distinction because the two parts are almost inseparably related – that is, in fact, the core idea of Korsgaard's constitutivism. But I hope the merits of separating questions around agency from the more directly normative parts of the theory will prove themselves over the course of this thesis. To reiterate, this chapter is meant to be explorative, meaning that I will try to reconstruct the key elements of Korsgaard's work without taking too much of a normative or argumentative stance.

The two parts I propose to distinguish between will be the topic of discussion shortly. To better understand them and their significance, it is important to first briefly point out that Korsgaard at times frames her theory as an attempt to answer to skeptical challenges,⁹ or to counter realism.¹⁰ The goal of her constitutivism may be stated as providing substantive and, presumably, truth-apt answers to moral questions while steering away from realism, which Korsgaard deems problematic.¹¹ I will leave it to the reader's discretion to establish precisely where this puts Korsgaard's theory in the meta-ethical landscape and hope that these brief remarks help gain a better preliminary understanding of my description and discussion of her theory and its focus on agency and its constitutive standards.

⁵ SC xi; CV 99

⁶ SC xii. Both 'making' and 'ourselves' deserve emphasis.

⁷ SC 214.

⁸ Korsgaard herself proposes a distinction between the 'constitution requirement' and the 'self-constitution requirement' (CV 99); Luca Ferrero divides constitutivism more generally into the claims that (i) morality can be derived from the constitutive principles of agency and that (ii) agency is necessary (Luca Ferrero, "Constitutivism and the Inescapability of Agency," in Russ Shafer-Landau, *Oxford Studies in Metaethics* 4 (Oxford: Clarendon Press, 2009), 303-333, 304); Stefano Bertea, whose distinction I will broadly follow, argues that constitutivism can be divided into conceptual analysis and normative principles (Stefano Bertea, "Constitutivism and Normativity: A Qualified Defence," *Philosophical Explorations* 16, no. 1 (March 2013): 81-95).

⁹ David Enoch was right to point this out (David Enoch, "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action," *The Philosophical Review* 115, no. 2 (2006): 169-198, 172f), although in her later work there are few traces left of this emphasis on addressing the skeptic (exceptions can be found in SC 29 and 32, for example).

¹⁰ See, for example, SC 112; SN 33; SN 34; SN 40.

¹¹ For her arguments against realism, see SN 1.4.1-1.4.9.

1.1 Action, agency and self-constitution

We human beings are forced to make something of ourselves and our lives. Such is the human condition, according to Korsgaard, and that analysis is the cornerstone of her theory, as she writes that human beings are ‘*condemned to choice and action*’.¹² The point, clearly, is that human beings *must* do something, because the opposite is impossible: only by dying, falling asleep and some other exceptional circumstances can we be relieved from the burden of choice and action. Even doing nothing is the result of some form of choice and as such even standing still and refusing to move are things we do.¹³ Action, in short, is the ‘simple inexorable fact of the human condition’.¹⁴

An equally crucial ‘fact of the human condition’ for Korsgaard is that human beings are reflective beings, or at least have reflective capacities. Although this is an important step, it is largely left out in the opening chapters of *Self-Constitution*. But in other work Korsgaard treads more carefully and argues that human beings are ‘self-conscious rational animals, capable of reflection about what we ought to believe and to do’.¹⁵ Evidently, the fact that human beings are, or can be, self-conscious has profound consequences for the way they must handle their human condition of being ‘condemned’ to choice and action. Self-consciousness allows us to take a step back and ask ‘why’. When we encounter in ourselves a strong impulse or desire to act, we have become aware of that desire and as a result it loses some of its initial force. Through self-consciousness and the awareness of our impulses that follows from it we can ask whether any particular desire really provides us with a *reason* to act.¹⁶ As a result, ‘[e]ach impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action’.¹⁷ The reflective distance we, qua reflective beings, have to our impulses means that we have the ability not to be impelled into doing or choosing something by desire. For Korsgaard,

it is as if there were something over and above all of your desires, something that is *you*, and that decides which if any of your desires to gratify. You may of course *choose* to act on a desire, but that does not mean that you are impelled by it. It means you take this desire as a reason, or, in Kant’s language, you make it your maxim to satisfy this desire.¹⁸

¹² SC 1. Italics in original.

¹³ SC 1

¹⁴ SC 2.

¹⁵ SN 46; Cf. Christine M. Korsgaard, “From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action,” in *Kant on Emotion and Value*, ed. Alix Cohen (Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan, 2014), 33-68, 41.

¹⁶ SN 93.

¹⁷ SN 91.

¹⁸ KE 57. Italics in original. Note that while Korsgaard here permits herself to either pretend or assume that there is something ‘over and above all of your desires’, in later work she argues that that is wrong: our identity arises only through the process of endorsement and therefore does not exist prior to choice and action (SC 19). What the overarching ‘you’ is in this passage, then, is an open question.

In this somewhat loosely formulated passage, Korsgaard makes it clear that because of our reflective nature, no desire has normative force on its own.¹⁹ Every desire needs to pass the *reflective endorsement test* – the test moral agents use to establish which of their inclinations and motives, if any, they will accept as being normative for them.²⁰

There are, based on the point above, two types of behavior. First, there is behavior that is caused by things working *on* or *in* the person in question. This category includes such things as twitching an eyebrow, but also a slip of the tongue.²¹ It comprises all behavior that is not caused by the person expressing it. The second category includes all behavior, by contrast, that *is* caused by that particular person. Korsgaard states that this type of behavior – ‘actions’ – must be expressive of oneself, resulting from one’s ‘entire nature working as an integrated whole’.²² How that works should become clearer in the remainder of this chapter and in chapter 3, but the important point for now is that behavior, for Korsgaard, can only be expressive of ourselves if it is based on reasons. And these need to withstand reflective scrutiny if they are to count as reasons at all – anything else would, after all, likely contradict Korsgaard’s non-realist aspirations. So how does reflective scrutiny turn something into a reason?

Reflective endorsement

The reflective endorsement test starts as follows. If we want to establish whether we have a reason to do or choose X, we must regard it as good. Trying to avoid realism, Korsgaard argues that whatever value X has does not lie in X itself, but stems from our desires and inclinations.²³ Hence, X is important to A simply because A takes X to be important to himself – not because X is somehow intrinsically valuable. As a result, A must regard himself as important in some way, because otherwise he could not transmit value to X. Thus, if we want to value something or have a reason to do something, we must value ourselves qua human beings and therefore ‘the value of humanity itself is implicit in every human choice’.²⁴ At this point, it seems as though Korsgaard puts an arbitrary stop to the regress of value. But, as she states, she draws heavily on Kant’s work,²⁵ meaning here that further regress is non-arbitrarily prevented by reference to the idea that human beings and all rational beings in general have what

¹⁹ In the next chapter I will argue that this reflective nature is in fact inescapable and can serve to ground agency.

²⁰ SN 89. This test will be the topic of the next section.

²¹ SC 18.

²² SC 19.

²³ SN 122.

²⁴ Ibid.

²⁵ Ibid. As Sem de Maagt rightly points out, constitutivists such as Korsgaard walk a fine line between Kant scholarship and Kantian ethics (“Constructing Morality: Transcendental Arguments in Ethics” (PhD diss., Utrecht University, 2017), 260f). Here and elsewhere I will use this to provide a generous reading of Korsgaard’s work, assuming that her work can be supplemented with Kant’s work if there is sufficient textual justification to do so. This, of course, requires to read Kant in a Korsgaardian way, which is just one of many ways of interpreting his work (for one thing, Korsgaard makes a controversial interpretative distinction between the CI and the moral law (see also footnote 153)). It also requires to read Korsgaard in a Kantian way which I believe to be the correct but perhaps not the only possible interpretation.

Kant calls 'absolute worth'.²⁶ Again, this absolute value is not realistically given, but is, according to Kant, the way rational beings necessarily represent themselves.²⁷ If they did not, all value would be conditional and contingent, ruling out the possibility of supreme, objective practical principles.²⁸ The reflective endorsement test can thus be summarized as follows:

1. To have a reason to do X, you must regard doing X as good.
2. You cannot regard doing X as good in itself, but can only regard doing X as good because it satisfies your needs, desires, inclinations, etc.
3. You cannot regard your desiring or needing to do X as making it good unless you regard yourself as valuable.
4. Therefore, you must regard yourself as valuable, if you are to make any rational choice.²⁹

If my interpretation is correct, Korsgaard advocates the view that this procedure of the transference of value also applies outside of the moral domain. Suppose that, while writing this thesis, I get a headache which I wish to get rid of, so I decide to get an aspirin. It might seem like too much of a stretch to argue that one would need to value one's existence or one's human nature in order to value the aspirin enough to take it. But according to the Korsgaardian analysis of value, it is not, because to have a reason to take an aspirin, I would have to think that an aspirin will reduce my headache and that reducing my headache is good. And reducing one's headache is not intrinsically good: it is only good because one has a desire to get rid of it. And such a desire, in turn, is valuable only because one regards oneself as valuable. There is simply no reason to improve your well-being – to try to reduce your headache – if you do not value yourself.

Practical identity and reasons

Still, the procedure described above is incomplete, because it lacks an important element of Korsgaard's work: practical identity. She defines that as a 'description under which you value yourself and find your life worth living and your actions worth undertaking'.³⁰ One's identity, for Korsgaard, thus is not a value-neutral thing. By contrast, from the quote it should be clear that she thinks identities are crucial to one's reflective process precisely because they are not value-neutral. Korsgaard insists that our conceptions of our practical identities form the basis of choice,³¹ because they provide people

²⁶ G4:427-429. See further CPR 5:87, for example.

²⁷ G4:428-429. The argument is of course debatable, but because my goal here is mostly descriptive, I will grant it for the sake of brevity and clarity.

²⁸ G4:428.

²⁹ The summary comes from Robert A. Stern's "Transcendental Arguments," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2019, 5. 'Doing' here might be substituted for 'choosing'.

³⁰ SC 20; SN 101.

³¹ SC 20.

with potential reasons for action.³² Consequently, practical identities must feature in the argument more prominently than they have done in the reconstruction above.³³

One could argue that the more personal, particular practical identities such as being a husband, being a Liverpool fan et cetera need not feature in the argument: it might suffice to include only the most basic practical identity – rational nature.³⁴ Yet for Korsgaard, that is not enough. She argues that, even though practical identities need not always feature in our reasoning, without them people would no longer have reasons to act.³⁵ Since practical identities ‘confer a kind of value on certain whole actions’,³⁶ they must feature somewhere in the argument, and the second step seems the most suitable place for them. Korsgaard seems to also want to include needs, desires and inclinations in the deliberative process, so the second step would have to include those *and* practical identities, especially because they cannot be reduced to one another – desires do not stem from practical identities and vice versa. Practical identities, to return to them more directly, give rise to reasons. Being a Liverpool fan, for instance, gives me a reason to watch their games, whereas being a husband gives me a reason to stop watching Liverpool at least for a moment when my wife comes home. The trick for any human being then is to unite one’s practical identities into a coherent whole.³⁷ This, presumably, means trying to strike a balance between the different identities so that they can co-exist, receive due attention and point one to what one has most reason to do. Building upon the version from the previous section and (I believe) bringing that closer to Korsgaard’s work, the reflective endorsement test would thus be as follows:

1. To have a reason to do X, you must regard doing X as good.
2. You cannot regard doing X as good in itself, but can only regard doing X as good because the relevant practical identity you value most in the given circumstances gives you reason to do X or because X satisfies your needs, desires, inclinations, etc.³⁸
3. You cannot regard your practical identity or your desiring or needing to do X as making it good unless you regard yourself as valuable.
4. Therefore, you must regard yourself as valuable, if you are to make any rational choice.³⁹

³² I will return to this point in more detail later on.

³³ Stern also provides a second, more elaborate reconstruction of Korsgaard’s argument, but permits himself to do so with hardly any references to Korsgaard’s writings.

³⁴ Korsgaard hints at the idea that rational nature is the most basic practical identity in, for example, SN 115; SN 120-122.

³⁵ SC 21; SC 23; SN 120-121.

³⁶ SC 21-22. A more precise way of putting it at this point would have been that practical identities ‘confer value on choice- or action-options’, since actions, in this idealized model, come only after deliberation. It makes little sense to ask what reason you have to carry out a certain action if you have already done it (and if value is conferred only after the fact, then one could ask why you chose the action in the first place).

³⁷ SC 21.

³⁸ It seems to me that Korsgaard is reluctant to reduce desires and needs to any practical identity other than, perhaps, one’s practical identity as a human being. Given her earlier assertion that no desire has normative force of its own, she seems to have to reduce them to *something*, but since she does not make it entirely clear what that is, I here continue to distinguish between practical identities and needs or desires for the sake of clarity.

³⁹ See also SC 24.

Including practical identities in step 3 is important here because it shows the status Korsgaard ascribes to them: as with anything in the first steps, practical identities are not intrinsically valuable. Their value stems from our valuing them. In fact, one could argue that in the constitutivist framework, practical identities do not exist outside of our actions – they are constituted by our actions and thus cannot function as a realistically existing reason-giving entity. It is by acting on the basis of the reasons my wanting to be a good husband gives me that I constitute myself as a husband and can truly be said to be a husband.⁴⁰

One important point to add is that Korsgaard's claim is not that the transference of value makes the object or action in question objectively or intersubjectively valuable. One could be quite bad at drawing and still regard it as a valuable and enjoyable activity, even knowing that one's drawings will not make anyone else's life any better. The value of drawing comes from, and does not extend beyond, either the enjoyment one takes out of the activity of drawing or one's appreciation of one's practical identity of being an amateur artist. Whether anyone else has reason to value the perhaps unimpressive drawings is something this other person must establish. The drawings, again, have not become objectively valuable because they are valuable to one particular person.

To return to Korsgaard, she summarizes her points about the transference of value as follows:

Since you are human you *must* take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reason and your humanity is the source of your reasons, you must value your own humanity if you are to act at all.⁴¹

One thing that is perhaps in most need of clarification is the transference of value. At times, Korsgaard makes it seem as though the mere fact that X satisfies a desire we value suffices for us to be able to value X. But the issue is more complex. She should, I propose, be read as advocating the view that transference of value is subject to principles. We cannot just value any desire or potential reason stemming from a practical identity. If one wants to establish whether something can be a reason, one needs to see if the maxim of acting on it can be willed as a law by a being with the identity in question.

⁴⁰ Cf. SC 214.

⁴¹ SN 123. Korsgaard might be accused of smuggling in a realism about the value of humanity here. But the reason for valuing humanity might be simply practical: we need to act, and in order to act, we need to value something. Korsgaard would then still have to show that 'humanity' is not an arbitrary thing to value if one wants to stop a regress of value, but that might well be possible without resorting to realism. One way of doing so – that is, one way of explaining why humanity is the source of value – is by arguing, as she does, that rational beings must regard themselves as free and that they see their ends as valuable because they have freely (autonomously) chosen them. Rational autonomy ('humanity' in Kantian terms) thus must be regarded as valuable and a source of value, because there is nothing beyond that nature that could transfer value for rational, autonomous beings (see KE 239-241). Alternatively, Korsgaard may be read as advocating the view that by valuing one's practical identity because it gives one reasons for action, one already values oneself as a being that needs those (reasons for action) and thus one already values oneself as a rational, reflective being and thus one also already values one's humanity (SC 22-25).

There are two possible answers to such questions, and both are normative: if it can be willed as a law, then it is a reason and then it is normative. If it cannot be willed as a law, then it must be rejected and that gives rise to (moral) obligation.⁴² So if we find in ourselves a desire, we, qua reflective beings, must ask if we can take that desire to be a reason for action and thus see if we can will the corresponding maxim as a law. If we can, then the maxim that describes our corresponding action, in the end, thus expresses our conception of a law.⁴³ To conclude, the entire process of reflective endorsement establishes what counts as a reason and thus it may be summarized as follows: “‘Reason’ means reflective success’.⁴⁴ Just how one may go about establishing whether something can be willed as a law and thus whether it can be a reason is the topic of the next section.

Reflective endorsement narrowed down: contradictions in the will

The previous section purposefully left underspecified why ‘being able to take a desire to be a reason’ must be the same as ‘being able to will it as a law’. This is not an obvious step, and Korsgaard has been criticized for making it.⁴⁵ This current section will be devoted to the question why and how Korsgaard equates willing something as a reason with willing it as a law, as well as to the distinct but related question how one goes about establishing whether something can be willed as a law.

For Korsgaard, it is essential that an action be attributable to an agent.⁴⁶ This requires that actions are seen as a combination of an act and an end,⁴⁷ of which the agent can see herself as the cause. Only such a combination is to be called ‘action’, which is therefore a technical term. The reintroduction of the agent here is crucial, and brings us back to constitutivism about agency. If an agent is to be able to truly see herself as the cause of an action, it must be expressive of herself and her will.⁴⁸ The action must therefore include not only the act, but also the end the agent aims to achieve. It must therefore be described not as only an act or only a reason, but a combination of the two, a maxim – the subjective principle on which people act.⁴⁹ These take the form of ‘I will do action-A in order to achieve purpose-P’.⁵⁰

With the need for maxims in place, the next question is why they must be willed as a law and, equally important, what guides the process of establishing whether or not you can will these maxims (rather than only one’s ends) as a law.⁵¹ There are two ways of understanding Korsgaard’s answer to the first question.⁵² First, she may be read as accepting the Kantian view that the will is a causality and that it therefore, like all causalities, operates on the basis of laws. She invites this reading when she claims that reflective beings like ourselves constitute themselves as causalities when they deliberate,

⁴² See SN 113.

⁴³ KE 57.

⁴⁴ SN 97. Korsgaard here does not mean the mental faculty.

⁴⁵ See, for example, William FitzPatrick, “How not”.

⁴⁶ SC 100.

⁴⁷ SC 82.

⁴⁸ KE 57; SN 100; SC 18-19. Cf. CPR 5:99-100, for example.

⁴⁹ G4:401f; G4:421f. Cf. SC 10.

⁵⁰ KE 57-58; CA 218.

⁵¹ I will be using these terms somewhat interchangeably, but the difference matters.

⁵² I thank Micha Werner for pointing this out.

for example.⁵³ Second, she may be read as proposing that we must turn our maxims into universal laws or principles because we must constitute ourselves as unified agents, meaning that a law we make for ourselves at t1 must be one we can also will to act on again at t2.⁵⁴ However, the second interpretation seems in fact to be saying the same as the first, albeit in different terms and in a more developed way. The first says that the will must operate on the basis of laws, the assumption being that this enables the will to truly be a causality by itself (that is, to be autonomous). The second says that one must act on the basis of laws because otherwise one fails to be unified, and this claim comes only *after* Korsgaard has explained why agential unity and autonomy (and *a fortiori* being a causality) mean the same.⁵⁵ While she seems to invite the interpretation that willing reasons as laws is only a conditional requirement, relying fully on one's commitment to being unified, I believe Korsgaard on both occasions makes the same (Kantian) point, even though the phrasing differs.

Having established *why*, according to Korsgaard, reasons must be potential universal laws, the next question is *how* they come to be such. The most straightforward way of making out whether a maxim is fit to be a law is to take Kant's universalizability requirement and to interpret that as a test for maxims, a reading Kant himself seems to invite:

Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law*. Here mere conformity to law as such, without having as its basis some law determined for certain actions, is what serves the will as its principle, and must so serve it, if duty is not to be everywhere an empty delusion and a chimerical concept.⁵⁶

Hence, the most fundamental command for the will is that it conforms its maxims to 'law as such' and thus to *lawlikeness*. Korsgaard therefore interprets such passages as the one I have just cited as meaning that universalizability is a test for maxims.⁵⁷ Thus, the starting point is this:

Since the will is practical reason, and since everyone must arrive at the same conclusions in matters of duty, it cannot be the case that what you are able to will is a matter of personal taste, or relative to your individual desires. Rather, the question of what you can will is a question of what you can will *without contradiction*.⁵⁸

⁵³ SC 72.

⁵⁴ SC 202. Taken in isolation, this second reading clearly makes the requirement of universalizing one's reasons dependent on one's prior commitment to being a unified agent.

⁵⁵ I will return to these claims on several occasions.

⁵⁶ G4:402. Italics in original.

⁵⁷ KE 63.

⁵⁸ KE 77. Italics in original.

Although this quote at first sight seems to suggest the opposite, Korsgaard here and elsewhere takes the familiar Kantian position that the will must be autonomous and thus that it must not be shaped by something external to it. If it *is* shaped by something external, then it loses its status as a free will. Even more strongly put, it *cannot* even be shaped by something external to it, because of our reflective nature. Korsgaard's framework thus distinguishes between behavior that either does or does not originate from our will, and only the former can be actions in the technical sense.⁵⁹ We cannot *act* (carry out actions) on the basis of external influences, because our reflexivity entails that we act for reasons and those lend their normativity to something internal to us: the reflective endorsement test under discussion here.⁶⁰ And thus, bearing in mind the importance of agential unity,⁶¹ the only question is what we can will without contradicting ourselves.

There are three possible types of contradiction of the will, each with its own test. The first is the *Logical Contradiction* test: 'On this interpretation', Korsgaard writes, 'there is something like a logical impossibility in the universalization of the maxim, or in the system of nature in which the maxim is a natural law: if the maxim were universalized, the action or policy that it proposes would be inconceivable'.⁶² One of the standard examples is the maxim involving a false promise: if one were to universalize that maxim into a universal law, it would undercut the entire practice of making and believing promises, which would make promises impossible and thus make the maxim incomprehensible.⁶³

The second type is the *Teleological Contradiction*. This rules out contradictions in willing a maxim as a law for a teleological system of nature: it forbids (acting on) maxims that would lead you to act 'against some natural purpose', or that 'could not be a teleological law. The maxim is inconsistent with a systematic harmony of purposes, or with the principle that any organ, instinct, or action-type has a natural purpose for which it must be the one best suited'.⁶⁴ The idea here is that a maxim is ruled out if it contradicts either said 'systematic harmony of purposes' or something's own natural purpose.⁶⁵

The third type is the *Practical Contradiction*. There, the contradiction is that 'your maxim would be self-defeating if universalized: your action would become ineffectual for the achievement of your purpose if everyone (tried to) use it for that purpose. Since you propose to use that action for that purpose at the same time as you propose to universalize the maxim, you in effect will the thwarting of your own purpose'.⁶⁶ One example would be the maxim of playing tennis at times when the courts are

⁵⁹ Cf. SC 8-14. It is important to note that Korsgaard does not adhere to perfectionism about the will – not every action is perfect and the will is not always perfectly autonomous (SC 25; SC 45).

⁶⁰ The following types of contradictions, to clarify, should be understood as a part of the reflective endorsement test, the outcome of which, as said, is a reason. This process serves to specify what one can or cannot value as a reason, because the more general process of value-transference described in the previous sections might allow one to value moral and immoral acts alike. The contradiction tests here, however, are not meant themselves meant to describe any transference of value.

⁶¹ I will return to the importance of agential unity in chapter 3.

⁶² KE 78.

⁶³ Cf. Korsgaard's introduction to Kant's *Groundwork*, xix.

⁶⁴ KE 78.

⁶⁵ This might be seen as one the reasons why Kant claimed that suicide is immoral (see G4:422).

⁶⁶ KE 78.

usually empty. If you were to universalize that, so that everyone playing tennis would do so at times when the courts are usually empty, and you would simultaneously act on it, then you would, as Korsgaard writes, will the thwarting of your own purpose.⁶⁷

These three tests help establish what maxims one can and cannot will as a law. In other words, they help discover what reasons one has to do something and what obligations one has not to do something. To summarize, then, Korsgaard's theory has thus far been shown to rely on the idea that human beings are reflective beings who therefore need reasons to act. To see what counts as a reason, they need the reflective endorsement test. This involves an appeal to their practical identities, desires, et cetera and it involves the fact that they must in the end value their humanity to ground the whole evaluative process. Furthermore, if the action that is to be based on the reason that is the outcome of the deliberative process is to count as truly one's own action, then it must be caused by oneself and not lead to contradictions within one's will. The contradiction tests help see what maxims meet Kant's requirement for maxims, which is '[nothing but] to conform [to] the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary'.⁶⁸

At the beginning of this section I claimed that the agent had come back into the picture and that this return was crucial, because Korsgaard's theory describes the constitution of agency, not of action. Yet now it seems the agent has disappeared off the stage once again, because most of what I have been discussing revolved around action and reflection. If Korsgaard is right, however, then the agent has anything but disappeared. In the next section I will discuss why, according to Korsgaard's theory, nothing has been more prominent here than agency.

Agency and its constitutive principles

In the previous sections I have discussed Korsgaard's analysis of action and her ideas on how action relates to our human and reflective nature. The discussion has almost exclusively focused on reasons and action, so why is her most recent book on the matter called *Self-Constitution: Agency, Identity and Integrity* and not *Action-constitution*? The basic answer is simple: action has a function, and it is to constitute an agent.⁶⁹ All the preceding work therefore is a groundwork for the (self-)constitution of agency. To see why, we need to look at Korsgaard's analysis of agency, which is the topic of this section.

'To be an agent,' Korsgaard writes, 'is to be the autonomous and efficacious cause of events in the world'.⁷⁰ To glance forward at where the argument is going, I will quote the next two sentences as well: 'In order to be efficacious, you must conform to the hypothetical imperative, and in order to be autonomous, you must conform to the categorical imperative. By conforming to these principles [...]

⁶⁷ The example comes from Barbara Herman's *The Practice of Moral Judgment* (Cambridge, Mass.: Harvard University Press, 1996), 138-141. It could be interpreted as an example to illustrate the Practical Contradiction, but also (in an entirely different context) as a false negative, as Allen Wood does (*Kant's Ethical Thought* (Cambridge; New York: Cambridge University Press, 1999), 105).

⁶⁸ G4:421.

⁶⁹ SC 82. Korsgaard need not be committed to the claim that agential self-constitution is the only function of action, but it is the one she seems most interested in.

⁷⁰ CV 99. See also CA 13; SC 82-83.

you make yourself into an agent: you *constitute* yourself as an agent'.⁷¹ Korsgaard thus lets there be no misunderstanding about the constitutive principles of agency: autonomy and efficacy. Just what they entail will be the topic of the next two subsections, but I will first briefly discuss why they are the principles of agency to begin with. Korsgaard puts the point clearly in the opening sections of *The Constitution of Agency*. She begins by wondering what makes a self-determining agent.⁷² In Kant's work she finds the answer: a self-determining agent must be supposed to have a free will – a will that is not determined by anything outside of itself and that can set itself its own principles.⁷³ Such a free will therefore must be an autonomous will, which is why autonomy is closely bound to the practical question of what we ought to do: 'to think [...] about what you ought to do is at the same time to think thoughts about what you would do were you a fully self-determining being'.⁷⁴ Again, autonomy will be discussed in more detail momentarily, but for now suffice it to say that someone can only be a self-determining agent if she has a free will.⁷⁵

The other constitutive requirement of agency, as said, is efficacy. It is constitutive of self-determining agency because without efficacy an agent would be unable to impact the world.⁷⁶ Agency would then be useless in the practical sense. A person might know perfectly well what she has reason to do or to refrain from, but if she is unable to do something with that knowledge – if she is not efficacious – she cannot, Korsgaard seems to claim, be called an agent. In the next subsections, I will respectively discuss the principles of autonomy and efficacy in more detail.

Autonomy

Korsgaard takes as a starting point Kant's claim that human beings, qua rational beings, must act 'under the idea of freedom' – they must regard themselves as free; they must at least act as if they have a choice.⁷⁷ That is, for Kant, reason 'must regard itself as the author of its principles independently of alien influences',⁷⁸ and the same goes for practical reason and thus the will.⁷⁹ The human will, Korsgaard seems to agree with Kant, 'can indeed be *affected* but not *determined* by impulses, and is therefore of itself [...] not pure but can still be determined to actions by pure will'.⁸⁰ It is therefore entirely up to the agent to determine what she should do. It is of course perfectly possible to subject ourselves to other people's judgments or to give in to desire, but agency and the nature of our will

⁷¹ CV 99. Italics in original.

⁷² CA 11.

⁷³ CA 12.

⁷⁴ Ibid. Cf. G4:454.

⁷⁵ Or if she can be supposed to have one, to take a more modest metaphysical stance.

⁷⁶ Cf. CA 13.

⁷⁷ G4:448; KE 162; SC 153. I will return to this in more detail later, but the point is that human beings might not know for sure that they are free (in fact, Kant seems to think that they cannot possibly know this), but they must at least act as if they know they are free. Freedom here is not dissimilar to courage, for example: one might not know beforehand whether one is courageous, but one may have to act courageously nonetheless and thus constitute oneself as a courageous person.

⁷⁸ G4:448; Cf. CPR 5:48.

⁷⁹ The will and practical reason, for Kant, are identical (see, for example, Pauline Kleingeld and Marcus Willaschek, "Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law," *Philosophers' Imprint* 19, no. 6 (2019), 4; G4:412).

⁸⁰ MM6:213.

require that we decide to do so ourselves: we must consciously decide to follow other people's judgment or try to satisfy a desire, for example. Only then will such things that are external to our will have normative force.

Korsgaard advocates the view that agency is self-determination.⁸¹ And self-determination requires that we 'act under the idea' that we can be self-determining – it requires that we regard ourselves as potentially self-determining, even though we might not know for sure that we are.⁸² It therefore follows that agents must regard themselves as autonomous and that they must regard the corresponding laws as imperatives for themselves.⁸³ There is an apparent paradox here: it seems contradictory to hold that agents must regard themselves as autonomous and that they must accept certain laws as normative for them. Korsgaard solves this paradox as follows. As we have seen, she maintains that we need reasons to act and that this ultimately means that we cast our acts and ends into maxims which take the *form* of a law, with no content external to the agent's will being imposed on them. The same idea applies here: our choice and will are only constrained by the fact that they need to be potential laws for us.⁸⁴ In the practical domain, the will must be regarded as autonomous and thus nothing outside of ourselves provides any content for it: '[A]ll that it has to be is a law'.⁸⁵ At this point, then, I can make explicit what has been implicitly present. The only principle under which an autonomous will acts is that it must choose a maxim it can regard as a law.⁸⁶ And that is precisely what Kant's Categorical Imperative (CI) prescribes: '*act only in accordance with that maxim through which you can at the same time will that it become a universal law*'.⁸⁷ Hence Korsgaard's conclusion that the categorical imperative *is* the law of an autonomous will.⁸⁸ The only 'constraint' for an autonomous will is that it must choose its own principles and set itself those principles as laws, and that is also what Kant's CI prescribes. It is a misunderstanding, therefore, to ask whether the CI can serve as a principle that would enable people to be autonomous; the CI is not a further principle that *enables* people to be autonomous. It is the very principle of autonomy itself – the principle of ruling over ourselves and thus giving laws to ourselves – and it is not an external principle an autonomous will can or cannot adopt.⁸⁹ Insofar as the will is autonomous, the CI already is its principle.

In sum, the argument is that when a rational being chooses a certain action by determining what she has reason to do, she shapes her will by conforming it with the principle of autonomy – the CI – and as a result she constitutes herself as a self-determining being and thus as an agent. She constitutes herself as an agent by deciding what she has reason to do and what she can will autonomously and without contradiction. By doing so, she establishes what she herself can autonomously will and cause, and that means that she establishes herself as an agent.

⁸¹ CA 33f.

⁸² I will return to this in 2.2.

⁸³ G4:454.

⁸⁴ KE 166.

⁸⁵ SN 98; KE 166.

⁸⁶ SN 98.

⁸⁷ G4:421. Italics in original.

⁸⁸ SN 98; KE 166; CA 110.

⁸⁹ CA 119-120; CA 62; SC 153.

Efficacy

As said before, the second constitutive principle of agency is efficacy. For Korsgaard, it arises out of the more general problems I have been describing above, combined with the absolute freedom that goes with autonomy. That is, the guiding question now is how human beings can use their autonomy to produce effects in the world and to do so in an effective manner. In other words: the question is how we are to ‘exercise our autonomy efficaciously in the world’.⁹⁰

An agent is efficacious, according to Korsgaard, if she ‘succeeds in bringing about whatever state of affairs she intended to bring about through her action’.⁹¹ The idea is that there are certain standards one cannot ignore in trying to bring about an action. For example, if you want succeed at dancing, you have to at least make efforts to not fall over.⁹² Trivial though that may seem, there is a more precise and interesting point underneath. Korsgaard argues that Kant’s hypothetical imperative dictates that whoever wants a certain end must also will the means to that end.⁹³ In Kant’s own words, ‘[hypothetical imperatives] represent the practical necessity of a possible action as a means to achieving something else that one wills’.⁹⁴ To return to Korsgaard’s example but now in the more precise way, if you want to succeed at dancing – if that is the end you desire – then the hypothetical imperative dictates that you take an effort not to fall over. Or rather, ‘not falling over’ does not seem like a means to the end of successful dancing, so the hypothetical imperative will more likely dictate that you practice a lot and then ‘not falling over’ will be a necessary by-product rather than an actual means to the end of successful dancing. But the general point stands. The hypothetical imperative allows us to use our autonomously chosen ends as input and then dictates that we should also will the means. Therefore the hypothetical imperative is constitutive of efficacy: by adhering to its requirements – by willing the means to ends you will – you become efficacious, because only by also willing and taking the means can one achieve one’s ends. And since efficacy is constitutive of self-determining agency, the hypothetical imperative is constitutive of agency.⁹⁵ To put the point in Korsgaard’s words, ‘[w]hen you act instrumentally [...], you *put* the belief that these are the means and your determination to realize the end together in the way required by the hypothetical imperative, and so *cause yourself* to act’.⁹⁶

Korsgaard, to conclude, tries to argue that the principles of autonomy and efficacy are not separate principles.⁹⁷ The hypothetical imperative, she believes, merely covers a particular aspect of the CI, namely that the laws of our will must be practical.⁹⁸ The point is this:

⁹⁰ CA 23.

⁹¹ SC 82. Korsgaard here invites the reading that one is only an agent when one actually succeeds in achieving one’s goals, but this is a stronger position than Korsgaard needs to commit herself to.

⁹² SC 82.

⁹³ Korsgaard’s introduction to Kant’s *Groundwork*, xvi.

⁹⁴ G4:414.

⁹⁵ Korsgaard discusses the importance of the hypothetical imperative for agency in more detail in “The Normativity of Instrumental Reason” (CA 27-68).

⁹⁶ SC 106. Italics in original.

⁹⁷ SC 70-72; SC 83-84; CA 67-68.

⁹⁸ SC 70.

[The categorical and hypothetical imperative] are so closely bound together that they seem to be inseparable, for nothing counts as trying to realize some end that is not also trying to determine *yourself* to realize that end, and nothing counts as determining yourself to realize the end that is not also trying to determine your own *causality*. In fact, [...] the two ideas are so closely linked that there is something artificial in the idea that there are *two* imperatives. There is really just one imperative here: act in accordance with a maxim you can will as a universal law. The hypothetical imperative merely specifies the kind of law we are looking for – a causal law, a practical law.⁹⁹

The argument would need more to get off the ground. Kant, for one, insists that the two imperatives are distinct, and for good reason: they differ with regards to normative status (one supposedly commands categorically while the other is normative only after someone sets himself a certain end) and their existence is shown by different types of evidence (one must be understood *a priori* whereas the other can supposedly be discovered empirically).¹⁰⁰ One could also argue that the supposed inseparability is only relative: if it works at all, then it only works from the categorical to the hypothetical imperative. The former, William FitzPatrick argues, certainly does not depend on the latter.¹⁰¹ Besides, as Korsgaard also notes, there are pragmatic reasons for keeping the two principles distinct, as I have been doing and will continue to do here.¹⁰² The two principles capture distinct features of agency and therefore it is at least more insightful to distinguish between them.

Agency revisited

Much of the previous discussion has been about action and deliberation, which are two things distinct from agency. But there is also crucial overlap. For Kant, as I have discussed, rational beings – agents, presumably – must operate ‘under the idea of freedom’ and thus view themselves as having a choice and, ultimately, as autonomous. Korsgaard taps into this conception of agency with her definition of agents being autonomous and efficacious. As a result, in order for human beings to count and continue to count as agents, they must act autonomously and efficaciously. Only then will they be agents.¹⁰³ And that is where the overlap with deliberation and action lies. The constitutive principles of agency are autonomy and efficacy, and actions are only really *actions* if they are based on a reflectively endorsed maxim. Since these maxims include an autonomously chosen end, and since the hypothetical imperative prescribes that willing an end requires that one also wills the means to that end, an action’s

⁹⁹ SC 81.

¹⁰⁰ See, for instance, G4:416 and 4:419.

¹⁰¹ See, for example, FitzPatrick, “How not,” 45. I will return to this point in 3.1.

¹⁰² SC 72. For some further elaboration on the distinction between the imperatives, see also CA 62f-63f.

¹⁰³ Korsgaard’s work is somewhat ambiguous about what this means for the stability of one’s identity and agency. On the one hand, she claims that identity and agency are constituted by actions and therefore do not exist independently, making it unclear how any sort of continuity and stability can be achieved (cf. SC 19). On the other hand, she claims that our identities *are* stable (SN 103). In the present context, however, it can remain unclear how stability is to be achieved. What matters most is Korsgaard’s constitutivism about agency and identity.

maxim conforms to categorical and hypothetical imperatives (which is tested, for example, by the universalizability tests) and thus agency and action perfectly match.

The overlap is, of course, no coincidence. It is the function of action to constitute an agent.¹⁰⁴ That is not to say that the constitution of agency is some covert goal that is implicit in every maxim. Agency-constitution is the result of autonomous and efficacious action and this is not implicit in the contents but in the *form* of a maxim.¹⁰⁵ A maxim's ends, we must autonomously choose because we are reflective beings and must operate under the idea of freedom; those ends' corresponding means, we must choose because we must be efficacious if we are to successfully deal with our human 'plight'.

With action, agency and agency's constitutive principles in place, Korsgaard distinguishes between two different requirements. First, there is the *constitution requirement*, which holds that 'unless the object conforms to the standard, at least to some extent, it ceases to be the kind of object that it is'.¹⁰⁶ This means that realism about agency is avoided by turning to constructivism (or constitutivism) about it: agency exists if and only if a person conforms to agency's constitutive standards. Agency only exists insofar as its constitutive standards are met and therefore it does not exist realistically, outside of, or regardless of, agency's constitutive standards. It also means that agency can exist in different degrees of success, as a process of falling apart and pulling yourself together.¹⁰⁷ This first requirement is met by all the points above, Korsgaard would likely argue, because agency is autonomous efficacy and without the constitutive standards – the categorical and hypothetical imperative – there would be no agency. Second, there is the *self-constitution requirement*, which holds that 'the object *makes itself* into the kind of object that it is *by* conforming to the standard'.¹⁰⁸ These requirements are requirements the constitutive principles of agency must meet.¹⁰⁹ Thus, autonomy and efficacy must enable human beings to make themselves into agents, friends, Liverpool fans et cetera.¹¹⁰ Suppose that I can reflectively decide that I have a reason to be a Liverpool fan, then the self-constitution requirement states that the hypothetical imperative must enable me to make myself into one. And it does, Korsgaard argues: we can take our practical identities and see what reasons or principles they provide us with, and then impose the principle of efficacy on them to discover what we ought to do in practice. And by adhering to these standards, we can in a somewhat paradoxical fashion make ourselves into what we are: by doing what the standards of being a Liverpool fan dictate, I can make myself into an actual Liverpool fan.

In sum, Korsgaardian constitutivism is the view that the combination of hypothetical and categorical imperatives checks the boxes for agency. They constitute agency, they allow rational beings

¹⁰⁴ SC 82.

¹⁰⁵ The form of a maxim, according to Korsgaard, is the relation between the action that is performed and the intention with which it is performed (KE 75-76).

¹⁰⁶ CV 99; CV 104.

¹⁰⁷ See SC 214. The claim that agency comes with varying degrees of success seems to entail that good and evil come in degrees too. Insofar as one takes Korsgaard to be trying to stay as close to Kant as possible, and insofar as Kant can truly be read as a no non-sense formalist, there might be an interesting topic for further research here.

¹⁰⁸ CV 99; CV 108.

¹⁰⁹ CV 108.

¹¹⁰ See, for example, CV 111.

to turn *themselves* into agents and they allow agents to maintain their agential status even though they may slip up at times.

1.2 Constitutivism and normativity

If agency and its principles are binding,¹¹¹ then according to Korsgaard this has important implications for morality and for ourselves. However, it might not be immediately clear how this connection between agency's constitutive principles and normativity is supposed to work.¹¹² It would be an oversimplification to state that agency's constitutive principles, especially the CI, are obviously normative. The danger here is that constitutivism makes normativity seem simpler than it is. It is a legitimate question to ask how the normativity works in Korsgaard's go-to example of building a house, and the same goes for the constitutive principles of agency, with the difference being that their normativity is even more in need of explanation than the normativity of house-building. Does it lie in contradiction, and is that somehow inherently morally wrong? Does it lie in our agential nature? And how could that ground normativity if agency arises only through action, and thus does not exist outside the domain of action and therefore cannot ground the normative side of that domain? In short, there is more explanatory work to be done, which will be the goal of this brief section.

A first way of grounding the normativity of agency's constitutive principles could be to, again, point to the value of humanity. If I value my reasons with reference to my rational nature, then non-contradiction requires that I grant that you can do the same and that therefore your reasons are as normative as mine. But they would not be normative for me – I could not be forced to make *your* humanity normative for me.¹¹³ Appealing to the CI as expressed in the Formula of Humanity here does not work. If I were to claim that I should take your humanity as normative for me because I ought to always respect humanity and thus never treat it as a mere means,¹¹⁴ then there would be clear moral duties for me; but this would presuppose that the CI is normative, and that normativity is precisely what is in question here.

In earlier work, Korsgaard seems happy to take a realist position about the value of humanity, writing explicitly that 'realism is right'.¹¹⁵ She thus contradicts her later views about the transference of value, where value exists only because the human will transfers value, meaning that all values and

¹¹¹ This is a big 'if', which is the topic of chapter 2.

¹¹² Cf. Berteau, "Constitutivism," 82.

¹¹³ SN 134. Korsgaard seems to disagree with Gewirth and his PGC on this point. Gewirth argues that agents must acknowledge that they have rights to freedom and well-being and that they must generalize this into also acknowledging that any other prospective agent has these rights. Gewirth states that agents must act in accordance with the 'PGC', the Principle of Generic Consistency: '*Act in accord with the generic rights of your recipients as well as of yourself*' (*Reason and Morality*, Paperback ed. (Chicago: Univ. of Chicago Press, 1995), 130-135). For Gewirth, then, agential self-understanding and consistency are such that they must immediately recognize other agents' rights (and value, supposedly), whereas Korsgaard thinks the argument needs more (as I show in the rest of the paragraph). (See also Gewirth's *The Community of Rights* (Chicago: University of Chicago Press, 1996), esp. 16-20.)

¹¹⁴ G4:429.

¹¹⁵ SN 166.

reasons are human creations.¹¹⁶ These values are either the conclusions of practical reasoning, which is not a description of a fact about the world,¹¹⁷ or they are ‘relational’,¹¹⁸ such that values arise only because we impose value on things, whereas realism would supposedly hold that X’s value exists irrespective of a human being’s valuing or devaluing of X.¹¹⁹ But this, of course, reintroduces the problem that also arose in my discussion of the reflective endorsement test. Why should we be able to impose value on things? That would, after all, require that we value ourselves too, and the question is what the status of that value is (if not realist) and where it comes from if we are to avoid realism about our own value. Korsgaard thus seems to somewhat help herself to an easy way out. As before, however, the answer lies in Kant’s work. To put the point very briefly, because I have covered it before, it works as follows. For Kant, human beings, qua rational beings, must operate under the idea of freedom. Hence, they must operate under the demands of autonomy and morality. Thus, we must impose value on our rational nature, if anything is to have value at all. And we must do so because the human condition is such that we must act, and action requires the process of reflective endorsement which, in turn, requires (and puts certain constraints on) the transference of value.¹²⁰ Therefore, realism about value is avoided – it is only imposed by human beings on practical grounds – and still humanity retains its ‘absolute worth’ and can thus serve to prevent the infinite regress of value that Korsgaard reinvents in her later work by trying to avoid realism. Without truly saying anything new here then, value-imposition is something rational beings necessarily do and this grounds the normativity of the process of value-ascription.

To repeat, valuing things as reasons for ourselves must be done autonomously. Consequently, heteronomy – roughly, the will’s being shaped by foreign impulses¹²¹ – is ruled out. In this Korsgaardian and perhaps Kantian framework, valuing and action arise solely from the autonomous will of an agent. Contradictions in the will are ruled out because they make such autonomous willing impossible. The wrongness of contradiction should therefore not be understood as being inherent to contradiction, but as being something which ultimately renders agency and action impossible. A further question then is whether agential unity or autonomy is not secretly taken to be inherently good. Korsgaard herself invites this question by claiming that, for Plato, a ‘just soul’ is ‘good to have for its own sake in virtue of its internal properties’,¹²² and by later taking Plato’s analysis of the soul as a model for her own work.¹²³ Yet she can easily overcome this problem of potentially smuggling in realism by virtue of her claim that we human beings are condemned to choice and action. Agential unity or autonomy is

¹¹⁶ SC 209.

¹¹⁷ CA 309.

¹¹⁸ SC 122.

¹¹⁹ SC 122-123; SC 112-113. Korsgaard points out several problems with realism, with one of her key objections being that realism cannot satisfactorily end the regress of normativity, because one could in principle always continue to ask what reason one has to do something. When asking why one should do X, it will not suffice to say that X is good, because then one could simply ask why one should do what is good, and so on, and Korsgaard seems to think that realism cannot deal with this regress (SN 32-33).

¹²⁰ Cf. SC 22-25.

¹²¹ G4:444; G4:441.

¹²² SN 110.

¹²³ See, for example, SC ch. 7.

not assumed to be good in itself, but because they are necessary (according to Korsgaard's analysis of human nature and action) for action, and action is necessary. Action and choice require autonomy and thus human beings should strive to be autonomous and thus constitute themselves as agents, because that is quite simply what the human condition requires.¹²⁴ Agential unity and autonomy are therefore not taken to be realistically valuable – something that is good to have in virtue of its internal properties, as Korsgaard phrases the realist position about agency – but are rather seen as a practical necessity and thus as things which only become valuable because human beings need them.

If my understanding of her work is correct, this is how Korsgaard grounds the normativity of the moral side of the constitutive principles of agency. The point is not that agency has constitutive principles and that these – especially the CI – are obviously normative. The point is rather that we are condemned to action and that our reflective nature requires that we be autonomous when choosing or acting. The CI then simply *is* the principle of autonomy and is therefore not inherently normative, but normative because of the particular kind of normativity generated by practical necessity and human nature: we need to act, and action requires autonomy, the principle of which (the CI) then gains its normativity on practical grounds. And, finally, the whole process of valuing choices and actions hinges on our necessarily valuing rational nature, rather than its being inherently valuable. Therefore, the same principles that have been argued to help us constitute ourselves as agents and that help constitute our practical identities are also shown to have normatively binding force in the moral general moral sense. Thus, the principles constitutive of agency, for Korsgaard, are the same as the principles of morality.

Earlier on in this section, I suggested that it is not clear how Korsgaard moves from valuing one's own humanity to valuing other people's humanity, and the same goes for their reasons for action. In *Self-Constitution*, Korsgaard makes a simple move to bridge the gap between one's own value and reasons and those of other people. She states, quite simply, that valuing other people and their reasons is the 'default position': ignoring other people's reasons or using them as a mere means to one's ends takes effort and constitutes a break with how people normally see one another.¹²⁵ As Korsgaard recognizes, it may be problematic to let other people's value and the normativity of their reasons depend on the way people 'normally' interact.¹²⁶ To overcome this, she returns to her earlier claims, some of which I have been discussing here and others which I will discuss later. She summarizes them as follows:

Without respect for the humanity in your own person, it is impossible to will the laws of your own causality, to make something of yourself, to be a person; and unless you make something of yourself, unless you constitute yourself as a person, it will be impossible for you to act at all.¹²⁷

¹²⁴ Cf. SC 213.

¹²⁵ SC 202.

¹²⁶ Ibid.

¹²⁷ SC 204.

The idea thus is that respecting one's own humanity is necessary. So is it not still possible to value only one's own humanity and disregard that of others? Korsgaard argues that it is not, because of the 'basic insight' that every person (which should here be understood as a unified agent) interacts with other people as he interacts with himself.¹²⁸ Although Korsgaard does not explicate this, I believe the idea is that it would be contradictory to make an exception of oneself and thus to value only oneself. And anyone who is successful enough at unifying herself will have avoided doing so and then it follows that she also values other people and their reasons.¹²⁹

In this chapter I have tried to provide a discussion and exposition of Korsgaard's writings and to present them as a unified theory of agential self-constitution. There are many possible points of critique, and these will be discussed in the next chapters. It is my belief, however, that a thorough, critical discussion starts with a fair and open-minded reading of the theory itself, and that is what I hope to have established in this current chapter.

¹²⁸ SC 206.

¹²⁹ Cf. SC chapter 9.

Chapter 2: The problem with agency

There has been an influential criticism of Korsgaard's theory, which seems to go right at the heart of the theory: Enoch's shmagency objection.¹³⁰ The problem, very briefly put, is that constitutivism fails to show that agency is necessary, or that it fails to show that it is necessary to understand what it means for someone to be an agent.¹³¹ In this chapter, I will take this critique as a starting point and assess several different ways of answering it.

The first part of this chapter will consist of a more in-depth explanation of the shmagency objection. I will argue that the problem cannot easily be dismissed. I will then provide a reading of Korsgaard's work on freedom as a counterargument against Enoch's critique. Then I will try to show why this counterargument does not work and give a first approximation of a different Korsgaardian answer to Enoch's challenge to her theory. As far as I am aware, she does not address Enoch's criticism in any explicit way, so I will construct an argument on the basis of her writings even though they are not specifically meant to refute Enoch's objection. I will argue that Korsgaard grounds the necessity of agency in human nature, which is why I will call it 'anthropological necessity'. After explaining this type of necessity, I will discuss two alternative readings of Enoch's shmagency, respectively interpreting shmagency as living a life in which one merely chases one's desires or in which one only half-heartedly engages in agency and action. Next, I will explore a different way of defending constitutivism, namely a line of argument proposed by Luca Ferrero, who argues that Enoch's point can be dealt with by showing that the whole practice of reason-giving (and thus also asking for a *reason* to be an agent) is part of agency.¹³² This approach, I will argue, leads to a trilemma. Having established that, the last section will comprise two things: first, an attempt to show that Korsgaard's theory can refute Enoch's critique and avoid Ferrero's (unintended) trilemma; second, a brief discussion of the general shmagency debate. This debate, I believe, is hindered by several philosophical errors or fallacies, such as attacking an unfair reconstruction of constitutivism or wrongfully shifting the burden of proof. Furthermore, the possibility of shmagency seems to be an assumption, rather than something that has been properly established. I will use this chapter to try to overcome this problem by discussing several potential types of shmagency. As I will try to show, none of them work, leaving the burden of proof with those who think that shmagency is in fact a genuine alternative to agency. Before I can really substantiate these claims, however, I will first dive into the debate.

2.1 Enoch's shmagency objection

For his first iteration of the shmagency objection Enoch distinguishes between a normative and metanormative level in constitutivism. At the normative level, he claims, constitutivists try to derive a

¹³⁰ David Enoch, "Agency"; Enoch, "Shmagency".

¹³¹ Enoch, "Agency," 179. Cf. Deryck Beyleveld, "Transcendental Arguments for a Categorical Imperative as Arguments from Agential Self-Understanding," in *Transcendental Arguments in Moral Theory*, ed. Jens Peter Brune, Robert Stern, and Micha H. Werner (Berlin, Boston: De Gruyter, 2017), 152.

¹³² See especially Ferrero, "Constitutivism".

‘full first-order normative [...] theory from what is constitutive of action’.¹³³ On the metanormative level, they also try to place second-order moral normativity in what is constitutive of action, so that moral normativity is on solid ground.¹³⁴ Enoch’s arguments focus mainly on this second-order normativity, and I will follow him in doing so.¹³⁵ If nothing can be derived from agency, then it is pointless to start a separate investigation into the possibility of deriving first-order claims from agency. And if only first-order claims can be derived from agency and they cannot be grounded in a way compatible with constitutivism, then the whole project is undermined as well.

The goals constitutivism as a general theory tries to achieve, according to Enoch, are as follows:

[Constitutivism] attempts [to account] for such (related) features of normative (or moral) discourse as its objective purport, its apparent universality (or claim thereto), its nonoptionality and nonarbitrariness, while nevertheless avoiding a more robustly realist metanormative theory [...].¹³⁶

With the discussion of Korsgaard’s theory from chapter 1 in mind, this is a broadly accurate description of the aims of constitutivism. The question Enoch then poses is whether grounding normativity in what is constitutive of action or agency can really account for all these theoretical goals.¹³⁷ His (negative) answer to this question then arises as follows. Suppose, Enoch starts, there is a moral skeptic who is willing to grant Korsgaard’s entire theory. He is willing to concede that Korsgaard is right in arguing that self-constitution is a constitutive aim of action and that the whole of morality and its normativity can be derived from the aim of self-constitution.¹³⁸ If this skeptic believes her entire theory, Enoch invites his reader to wonder, do we then have ‘any reason to believe that now [the skeptic] will care about the immorality or irrationality of his actions?’¹³⁹ Why, Enoch wonders, would it be impossible for the skeptic to respond along the following lines:

“Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can’t act without aiming at self-constitution, but why should I act? If your reasoning works, this just shows that I don’t care about agency and action. I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution.”¹⁴⁰

¹³³ Enoch, “Agency,” 170.

¹³⁴ Ibid.

¹³⁵ Ibid. Enoch raises interesting and insightful questions over constitutivism on the normative level as well. I will here leave these for what they are, because his points about constitutivism’s second-order aspirations seem to me more important and pressing.

¹³⁶ Enoch, “Agency,” 177.

¹³⁷ Ibid.

¹³⁸ Ibid., 179.

¹³⁹ Ibid.

¹⁴⁰ Ibid.

To bring the point home further, Enoch refers to an example Korsgaard uses on multiple occasions too: building a house.¹⁴¹ As with the agent, we could convince the builder that there are certain constitutive standards to building houses, such as providing shelter. Insofar as the builder is concerned with building a house, let alone a good one, he is bound by these standards. But why, Enoch asks, would he not be entitled to respond that he is not engaged in the activity of house-building?¹⁴² What if he does not care about house-building and its constitutive standards and is instead happy building a shmouse – something that might look like a house but does not live up to its standards?

The constitutivist, or the house-building fanatic, could reply to the skeptic or shoddy builder that he is not an agent, that he is not engaging in the game of agency or house-building if he does not live up to those games' respective constitutive standards. But the skeptic or shoddy builder could simply respond with 'so what?', Enoch claims.¹⁴³ The point, put more seriously, is that 'the status of being constitutive of agency does not suffice for a normatively non-arbitrary status'.¹⁴⁴ Thus, for Enoch the issue is that it may be granted that autonomy and efficacy are constitutive of agency just as 'providing shelter' is constitutive of being a house, but one could always ask why one should care about such principles. That is, it is possible to be uninterested in agency and if that is the case, agency's constitutive principles seem to have no normative force for you. Constitutivism would then, first, rely on some further argument to show that agency is necessary and, second, constitutivists would have to avoid a renewed shmagency objection, which calls into question why there would be a normative demand for agency if agency is absolutely necessary. Since the argument constitutivism needs for the necessity of agency would be external to agency – because its own demands do not suffice – constitutivism cannot possibly tell the whole story. Enoch takes the point further and argues that constitutivism would have to rely on realism to provide an outside reason as to why agency is necessary,¹⁴⁵ but the weaker point suffices to call constitutivism's ability to achieve its aforementioned goals into question.

To conclude, Enoch's point is that the constitutive standards of agency and action are only normative if you care about being an agent or if you care about performing actions, giving those standards relative, hypothetical force at best. So even if someone accepts Korsgaard's entire theory as it stands, Enoch argues that that person has no more reason to be an agent or to engage in actions than he had before accepting constitutivism. If that is correct, then constitutivism fails to ground normativity in the way it is supposed to. Self-constitution and indeed morality will then be up to everyone's own discretion, completely contingent on whether they care about said things or not. Enoch's conclusion can therefore be summarized thusly: constitutivism might be able to account for everything that follows once agency is established, but by itself constitutivism fails to establish the necessity of agency or action and therefore the entire project is contingent.

¹⁴¹ Ibid.; SC 28-30; SC 131; CA 8.

¹⁴² Enoch, "Agency," 179.

¹⁴³ Enoch, "Shmagency," 209.

¹⁴⁴ Ibid.

¹⁴⁵ Enoch, "Agency," 19; Enoch, "Shmagency," 209.

2.2 A Korsgaardian reply to Enoch

In response to Enoch's objection, Korsgaard can concede that constitutivism does indeed not directly establish the necessity of agency. One could wonder if constitutivism even really has to be able to do so, but there are two more interesting lines of thought to be found in Korsgaard's work. Even granting that Enoch is right in claiming that constitutivism cannot by itself establish the necessity of agency, Korsgaard has one argument that could be read as a response to Enoch's objection and, alternatively, her broader framework may ultimately prove to be able to show the necessity of agency, or so at least I will argue in the remainder of this chapter.

Enoch raises one objection to Korsgaard's constitutivist theory specifically. He restates his earlier argument about constitutivism's inability to clarify why agency is necessary or to ground its necessity simpliciter and adds that Korsgaard might be aware of the issue and that she deliberately leaves it unclear what kind of necessity underpins agency.¹⁴⁶ Supposedly, she thinks that action and agency are necessary, and elaborates on that only by saying that the necessity in question is not causal, rational or logical.¹⁴⁷ And she does indeed state that those three types of necessity are not the ones agency could be grounded on.¹⁴⁸ As Enoch also notes,¹⁴⁹ she claims that action is the human being's 'plight',¹⁵⁰ and Enoch suggests that that is all Korsgaard has to say.¹⁵¹ But there are more kinds of necessity she can invoke and more explanations she can turn to.

We human beings are condemned to action and choice, as I have discussed in the previous chapter. And that seems a fair starting point. Human beings, on Korsgaard's analysis of human nature, cannot do nothing. Although we can often refrain from doing something, Korsgaard suggests that we are still responsible for doing so: refraining from doing something is also something we *do*. Korsgaard's point appears to be that humans are reflective beings and must decide what to do, and this sets them apart from other things that can also be described in action-terms; trees are waving in the wind, waves are roaring and rivers are flowing down mountains, but none of them can be held responsible for what they 'do'. But assuming that it is true that human beings cannot do nothing, why would they have to engage in complex processes of reflection and why would they have to carry out *actions* – a certain class of acts which includes maxims, ends, the whole nine yards. Couldn't human beings in most cases just do what their desires tell them and be perfectly fine doing so? In Enoch's terms, couldn't human beings be 'perfectly happy performing shmactions – nonaction events that are very similar to actions but lack the aim [...] of self-constitution'?¹⁵² In fact, it seems more empirically plausible that most people often go about their lives rather unreflectively and that they are indeed 'perfectly happy' doing so. So how can Korsgaard go about showing that agency is in fact necessary? In this section I will interpret one of her essays as an answer to that question and in the next section I will put forth a different Korsgaardian answer.

¹⁴⁶ Enoch, "Agency," 188.

¹⁴⁷ Ibid., 188f-189f.

¹⁴⁸ SC 2.

¹⁴⁹ Enoch, "Agency," 189f.

¹⁵⁰ SC 2.

¹⁵¹ Enoch, "Agency," 188.

¹⁵² Ibid., 179.

Morality as a vocation

In “Morality as freedom”, Korsgaard discusses the way Kant connects freedom and morality. Although she does not address agency explicitly, I will argue that there is sufficient reason to read her argument as pertaining to agency as well. In this section, I will therefore go into her interpretation of Kant’s theory and try to provide the outlines of an argument for why this may be read as being a theory about both the source of moral normativity and the necessity of agency.

In order to arrive at the claim that morality and freedom are intimately related and that morality should take precedence over the non-moral inclinations that also form part of our human nature, Korsgaard puts forth two arguments: the Argument from Spontaneity (AS) and the Two-Worlds Argument (TWA). The former is meant to show why a free will makes the moral law (the CI¹⁵³) its principle; the latter why imperfectly rational beings ought to do the same.¹⁵⁴ The AS begins with the now familiar Kantian point that the will cannot be formed by something external to it.¹⁵⁵ As I have discussed in the first chapter, when reflective rational beings such as ourselves are confronted with a certain desire, they must decide whether or not they will act on it and must therefore form a maxim expressing their endorsement or rejection. The AS thus starts from the idea that human beings must act and choose as if they were free. That is not to say that they are in fact free or that they are a very special type of animal – one that mysteriously lifts itself above causal influences and becomes a causal force of its own.¹⁵⁶ The claim about the need to *regard* ourselves as free is in fact more modest. Korsgaard invites the reader to envision a case in which you participate in a scientific experiment.¹⁵⁷ An electronic device is put in your brain and that device programs your every move, not by bypassing your thinking processes but by taking control of them – it determines what you think and by doing so it also determines what you do. So suppose that you decide to spend your morning working. It may then occur to you that this was in fact determined by the device, and you instead decide to go shopping. But this, you realize, is also determined by the device. Yet although you can continue to second-guess the device, you will still need to do something. The point then is that even though the device has taken over your thought process, you will still need to think about what to do, just as you would without such a device. Consequently, Korsgaard concludes, it does not matter if we truly *are* free: what matters is that we must regard ourselves as being free and that ‘we must regard our decisions as springing ultimately from principles that we have chosen, and justifiable by those principles’.¹⁵⁸

To better understand the conclusions Korsgaard draws from her thought experiment, it helps to briefly liken it to Harry Frankfurt, who discusses whether people can be held morally responsible for

¹⁵³ Korsgaard in fact proposes a distinction between ‘the moral law’ and the CI, where the latter is the law of acting only on principles one could will as a universal law and the former is ‘the law of acting only on maxims that all rational beings could act on together in a workable cooperative system’ (SC 80; SN 99). Since she claims that Kant does not make this distinction (ibid.) and, more importantly, seems to use the terms interchangeably herself too, I will here not distinguish between them.

¹⁵⁴ KE 171.

¹⁵⁵ KE 162; G4:448; R6:24; Wood, *Kant’s Ethical Thought*, 51-52.

¹⁵⁶ Cf. KE 183.

¹⁵⁷ KE 162-163.

¹⁵⁸ KE 163.

something they have done even if they could not have acted otherwise.¹⁵⁹ Korsgaard argues that they can, but there is an important difference: while Frankfurt-style cases usually feature in metaphysical debates on the freedom of the will, Korsgaard states that she takes the Kantian approach, meaning that freedom is a ‘postulate of practical reason’.¹⁶⁰ Her idea, at least at this point in the argument, is that we must regard the will as free on practical and thus moral grounds, because belief in freedom is a prerequisite for ‘obedience to the moral law’.¹⁶¹ All that matters is that one regards oneself as free and the justification for doing so is practical. The question if the will actually *is* free can be thus bracketed.

With these modest ontological commitments in place, the next step is to define, as Kant does, the will as a ‘rational causality that is effective without being determined by an alien cause’.¹⁶² According to Kant, the will’s status as a causality entails that it – like any other causality – must operate on the basis of laws.¹⁶³ As discussed in the first chapter, both Kant and Korsgaard advocate the view that this ultimately means that the will needs a principle in order to be free and that this principle is the categorical imperative. Because this has already been the topic of the first chapter, I would here like to point to something related and of greater importance in the context of this current chapter: Kant’s definition of the will as an effective and autonomous causality is remarkably close to Korsgaard’s definition of agency as being the efficacious and autonomous cause of events in the world. Because of this overlap, I will treat what Korsgaard writes about the will as also applying to agency.¹⁶⁴

To return to the Argument from Spontaneity (AS), once one grants that the will needs principles or laws in order to be autonomous, the next step is that rational beings act on the basis of maxims. Thus, with the outlines from the first chapter in mind, the point is that we act on the basis of a subjective principle that describes the action and its end, such as ‘I will do this action, in order to get what I desire’.¹⁶⁵ On Korsgaard’s interpretation of Kant, however, this maxim only has normative force once you adopt a further maxim that states that you will make it your end to get what you desire.¹⁶⁶ But why should you adopt this underlying maxim? A regress of reasons seems inevitable. Thus if the CI is to be the ultimate principle of the will and if regress is to be avoided, then the CI must be shown to have some property that stops the regress.

¹⁵⁹ Harry G. Frankfurt, “Alternate Possibilities and Moral Responsibility,” *The Journal of Philosophy* 66, no. 23 (1969): 829-839, 829-930. See also his “Freedom of the Will and the Concept of a Person,” in *What Is a Person?*, ed. Michael F. Goodman, Contemporary Issues in Biomedicine, Ethics, and Society (Totowa, NJ: Humana Press, 1988), 127-144.

¹⁶⁰ KE 172.

¹⁶¹ KE 172; C2 5:132.

¹⁶² KE 163; G4:446; G4:453; G4:458; C2 5:15.

¹⁶³ G4:446. Cf. KE 163 and 167.

¹⁶⁴ In fact, it seems as if Korsgaard understands agency as meaning no more than ‘having a will’, where ‘will’ is of course understood in the technical, Kantian sense. I cannot here establish whether this is a fair description of her theory of agency and will thus stick only to the less controversial claim that her definition of agency and Kant’s definition of the will are remarkably similar (and sufficiently similar for my purposes here).

¹⁶⁵ KE 164.

¹⁶⁶ *Ibid.*

Korsgaard argues that the CI is in fact capable of stopping the regress of reasons. It does so by enabling one to choose one's fundamental principles.¹⁶⁷ Before the will can choose such principles, however, one first needs to find what things are true candidates to become reasons. In other words, the will and its CI can establish what lower-order principle takes precedence over other lower-order principles, but only after it is clear what the options are.¹⁶⁸ We may encounter situations in which different, conflicting maxims present themselves and the will needs to impose an order on them. For instance, we could be confronted with a maxim of what Korsgaard and Kant call 'self-love'¹⁶⁹ – which subordinates moral incentives to those of our inclination – and a moral maxim. These could be summarized as follows:

Maxim of self-love: I will do what I desire and I will do what is morally required if it doesn't interfere with my own happiness.

Moral maxim: I will do what is morally required and I will do what I desire if it doesn't interfere with my duty.¹⁷⁰

Thus far, these two maxims have been presented as genuine alternatives to one another. But they are not, Korsgaard insists. The object of discussion is a free will, and because this will is supposed to *remain* free, the moral maxim is the only real option.¹⁷¹ As seen before, the will is free if and only if it autonomously chooses its own principles. According to Korsgaard this means that the will thus loses its freedom once it chooses a maxim of inclination over a moral maxim.¹⁷² As Korsgaard puts it, '[t]he free will that puts inclination above morality sacrifices its freedom for nothing'.¹⁷³ That is to say, a free will choosing inclination above morality chooses against the maxim that will safeguard its freedom and therefore it gives up its freedom.¹⁷⁴ From the deliberative perspective of the will, then, the two maxims are certainly not on an equal footing.

The free and moral will is only one side of our human nature, however. As Korsgaard also acknowledges, human beings are much more than just a will: they are 'whole persons'.¹⁷⁵ It is quite natural for human beings to strive for happiness and it is to be expected that this striving will have the upper hand on more than a few occasions. Happiness certainly sounds more appealing than moral duty. Yet, Korsgaard warns, a will that adopts inclinations such as striving for happiness as its maxims subverts itself to natural, causal forces.¹⁷⁶ A person acting from self-love therefore is not (freely) acting at all, but is merely passively following his desires and inclinations. Even granting that that is true,

¹⁶⁷ See, for example, KE 165.

¹⁶⁸ KE 165.

¹⁶⁹ Ibid.; C2 5:22.

¹⁷⁰ See KE 165.

¹⁷¹ Ibid.

¹⁷² Ibid. It is an open question, as Micha Werner pointed out to me, if or how the person in question can be held accountable for doing this or letting this happen.

¹⁷³ KE 167.

¹⁷⁴ The point here is essentially the same as chapter 1's point that the CI is the principle of an autonomous will.

¹⁷⁵ KE 166.

¹⁷⁶ KE 168.

however, the Enochian questions return: why should someone care about that? Why should someone *act*? Isn't it perfectly possible (and probably more desirable) to leave duty for what it is and instead strive for happiness?

A human being's higher calling

The AS by itself cannot answer these questions, Korsgaard seems to acknowledge. 'What is needed', she writes, 'is an incentive for us to identify with the free and rational side of our nature'.¹⁷⁷ What is needed, thus, is an incentive or a *reason* for us to use or identify with our potential to be autonomous and efficacious, and thus we need a reason to be an agent.¹⁷⁸ Korsgaard thinks the Two-Worlds argument (TWA) can provide this reason. The TWA distinguishes between what Kant calls, among other things, a 'world of sense' and a 'world of understanding'.¹⁷⁹ Korsgaard calls these 'worlds' the 'phenomenal world' and 'noumenal world', respectively.¹⁸⁰ Before going into the argument, it is important to highlight Korsgaard's ontological modesty here. As with the AS, the point is not that these worlds actually exist – positing such a distinction would be what Hannah Arendt called the 'metaphysical fallacy' and an ontological dualism Korsgaard is in fact eager to argue against.¹⁸¹

The noumenal world is beyond the reach of our knowledge, Korsgaard states, meaning that it has the status of an assumption which serves to explain the character of the world we *can* know, the phenomenal world.¹⁸² To conceive of yourself as a member of the noumenal world means that you conceive yourself as providing some of the grounds of the world as we know it.¹⁸³ Korsgaard herself does not discuss in too much depth how that works or what it means 'to conceive yourself as a member of the noumenal world', but with Kant we find the idea that both 'worlds' are in fact standpoints from which we can regard ourselves and come to know the laws for our actions.¹⁸⁴ Our rational capacities enable us to see ourselves as members of a non-empirical world, independent of laws of nature and subject only to *a priori* laws of reason.¹⁸⁵ As rational beings, we can think about what we ought to do, even in the face of many conflicting and contradictory inclinations haunting us in the phenomenal world. To put the idea crudely, we have certain duties and obligations by virtue of being members of

¹⁷⁷ KE 167.

¹⁷⁸ Korsgaard explicitly claims that every action needs to involve an incentive and a principle, where 'incentive' means 'motivationally loaded representation' (SC 104-105). She here seems to disagree with Kant on the moral status and role of incentives, for Kant at times is quite critical of the role of incentives in morality (see, for example, G4:412-413) and proposes rather to look for a moral law that 'commands absolutely of itself and without any incentives' (G4:425). One could therefore wonder to what extent the role Korsgaard ascribes to incentives aligns with Kant's theory. Furthermore, one might ask how 'Kantian' it is to seek incentives or reasons for agency or morality (granting that one could resolve the potential contradiction in trying to reconcile the claim that action is necessary and that this somehow leads to agency with the idea that we might need a reason to be agents).

¹⁷⁹ G4:451.

¹⁸⁰ KE 168.

¹⁸¹ Hannah Arendt, *The Life of the Mind*, One-volume ed. (San Diego: Harcourt, Inc, 1981), 15, and KE 160, respectively.

¹⁸² KE 168.

¹⁸³ Ibid. Cf. G4:453.

¹⁸⁴ G4:452.

¹⁸⁵ Ibid.

the noumenal world, but we are easily distracted from them because we are also members of the phenomenal world. And by regarding ourselves as not only under the influence of inclinations and laws of nature but also as free members of the noumenal world, we can actively constitute ourselves, rather than being thrown around by desire.¹⁸⁶ As rational beings, we are potential members of the noumenal world and thus have a 'higher vocation', Korsgaard claims.¹⁸⁷

The combination of the TWA and the AS is meant to show why human beings ought to accept the moral law as their principle: it makes them autonomous and efficacious, Korsgaard finds. Thus it explicates, in effect, the reason why we should be agents. As rational agents, we are capable of engaging in something 'higher' than the 'mere' satisfaction of our desires and that higher (or highest) good is something to be strived for. Korsgaard herself summarizes the point as follows:

The Argument from Spontaneity shows why a free and spontaneous will, uninfluenced by anything, makes the moral law its principle. The Argument from the Two Worlds shows us why we imperfectly rational beings, influenced by sensibility as well as morality, should do so as well. If we are free we are members of the intelligible world, the ground of the sensible world and its laws. This gives us a "higher vocation" than the satisfaction of our own desires. We can help to bring about the Highest Good in the world. *The thought of that higher vocation is the motive of morality.*¹⁸⁸

Just why we ought to strive for the highest good is something that needs more explanatory work than Korsgaard gives it, and certainly more than I have given here. And there are further problems with this approach, but these will be the topic of the next section. For now, the important thing is that Korsgaard proposes, with Kant, that the noumenal and phenomenal worlds are in fact 'standpoints'¹⁸⁹ that can be broadly equated with practical and theoretical reason, respectively.¹⁹⁰ While from the latter perspective it seems perfectly understandable that a human being would disregard his duty in favor of some desire he has, from the deliberative, noumenal perspective that would be incomprehensible.¹⁹¹ For Korsgaard, the idea behind that – which she here does not make explicit – seems to be that the noumenal 'world' is subject to a priori laws of (practical) reason and it would be unintelligible for someone participating in this world to subject those laws to inclinations. With this in place, then, Korsgaard reformulates her argument for the 'possibility of morality' and, arguably, the necessity of agency as follows:

(i) We must act under the idea of (at least negative) freedom; (ii) we must therefore act on maxims we regard ourselves as having chosen; (iii) by the Argument from Spontaneity [...] we

¹⁸⁶ Cf. KE 169.

¹⁸⁷ KE 169; C2 5:88.

¹⁸⁸ KE 171. My emphasis. Korsgaard's point here brings her rather close to moral perfectionism. In chapter 3, I will argue that this does not apply to her theory outside of the particular paper under discussion here.

¹⁸⁹ KE 173. See also MM6:418.

¹⁹⁰ KE 174.

¹⁹¹ KE 173.

are led to the moral law (the positive conception of freedom); (iv) our ability to act on the moral law teaches us that we *are* (negatively) free; (v) if so, we are members of the intelligible world, and have a higher vocation than the satisfaction of our desires; and (vi) this provides us with the incentive to be positively free – that is, moral.¹⁹²

All this is meant to explain an earlier and crucial claim Korsgaard makes: we need an incentive or reason to ‘identify with the free and rational side of our nature’.¹⁹³ To return to Enoch, then, Korsgaard’s answer here seems to be that we naturally have the capacities that enable us to be free and efficacious and thus to be agents. And to answer to Enoch’s shmagent’s question, the point is that we have a strong motivation to be agents rather than shmagents or beings that follow their desires: morality will make us free. Korsgaard finishes her essay with a discussion of specific moral duties and virtues, and how these virtues will truly make one free. Rather than discussing that final section here, the important thing is that Korsgaard in this essay in general argues that morality will make us free and efficacious and thus that morality will make us agents. And she argues that we need a reason to behave morally and thus, *a fortiori*, to be agents, and that this reason is our supposed ‘higher vocation’: to live a moral life is better than to live a life centered around the satisfaction of desire. Or more specifically, it is better to live a life in which the precepts of morality take precedence over the satisfaction of desire whenever the two are in conflict. This approach, however, is unlikely to convince Enochian skeptics. In the next section, I will discuss why. I will there go into some of the problems with this Korsgaardian answer, before moving to what I will argue to be a more solid Korsgaardian answer to Enoch’s challenge, utilizing the elements from her work as described in the first chapter of this thesis.

The problems with Korsgaard’s answer

A first issue that needs to be dealt with has to do with my reconstruction of her essay as an answer to Enoch. In “Morality as freedom” Korsgaard attempts to explain the ‘possibility of morality’.¹⁹⁴ She does not discuss agency directly, so it might seem inappropriate to use that particular essay as an answer to the shmagency problem. However, as I have shown in the first chapter, Korsgaard defines agency as autonomous efficacy, and since these concepts *are* discussed in “Morality as freedom”, the essay may indeed be read as pertaining to agency as well. But even if her essay can indeed be read as an argument for the need to be agents, there are further, internal, issues with the argument. First, there are different ways of understanding the argument, and none is without shortcomings. The first possible interpretation I have outlined is the one reading the argument as stating that we need a reason to be moral and to be an agent. This reason then lies in the ‘higher vocation’ we have as beings with rational

¹⁹² KE 175. Italics in original. Cf. C2 5:86-87. I have left unspecified what Korsgaard means by ‘positive’ and ‘negative’ freedom, on the assumption that these terms are familiar enough. Korsgaard herself discusses these terms explicitly only in passing remarks as well (see, for example, KE 174 or 175). It is worth pointing out, however, that she claims that the idea of freedom and membership of the intelligible world leads to positive freedom because it provides a ‘conception of ourselves which motivates us to obey the moral law’ (KE 174). The final section of the essay might also be seen as a discussion of positive freedom (KE 176-183).

¹⁹³ KE 167.

¹⁹⁴ KE 175.

capacities: human beings are potential members of the noumenal world and therefore they are capable of engaging in something supposedly more worthy than ‘mere’ gratification. This allows Enoch’s shmagent – or the moral skeptic for that matter – to justifiably ask why he should engage in this supposed higher form of conduct. Agency’s call, on this interpretation, relies fully on one’s willingness to, first, grant that moral conduct is indeed ‘higher’ and, second, to actually have a desire to engage in such behavior, and that does not provide the agential necessity constitutivism is supposed to be able to deliver.

A more elaborate and perhaps slightly more convincing version of the first interpretation would state that human beings ought to behave as agents not only because that is the higher vocation they have qua beings with rational capacities, but also because they have a moral duty to do so.¹⁹⁵ This seems to be a possibility Korsgaard is inquiring into in the final section of her essay. The argument there is that human action is purposive – human beings set themselves ends.¹⁹⁶ And there are certain ends that are *duties* and thus we have a duty to have maxims of actions that promote those ends.¹⁹⁷ For example, Kant argues that human rational nature is an objective end.¹⁹⁸ As a result, we have a duty to respect the value of rational beings and this leads to some substantive duties such as a duty to ourselves to refrain from lying.¹⁹⁹ By setting themselves the maxim of refraining from lying, for instance, and by standing by that maxim – that is roughly what Kant calls ‘virtue’²⁰⁰ – human beings can control (not overcome) their inclinations and thus become free by being virtuous.²⁰¹

Even if the more elaborate version is correct, and thus if it is true that morality inherently imposes certain restraints on us in the form of objective ends and duties, the argument does not work against Enoch’s shmagent. As before, it is dependent on one’s prior commitment to morality. Even granting that it is true that there are certain moral duties that every rational being has, regardless of whether they actually accept them or not, such duties will only steer one into behaving morally if one is already willing to be moral and free. Both versions of the first interpretation of Korsgaard’s point outlined in the previous section thus presuppose, rather than show or argue for, a commitment to morality and agency.

However, there is another possible interpretation: Korsgaard’s solution may be read as an attempt to argue that the moral law is the only principle a free will can choose in order to remain free. The argument then is that we should let our rational, agential side take the upper hand in cases of conflict with inclination or desire because that is the only way to safeguard or constitute our will’s freedom. Again, however, this makes agency contingent on one’s wish to have a free will (and thus, according to Korsgaard’s Kantian line of thought, one’s wish to be autonomous and thus to be an agent)

¹⁹⁵ This might be thought to presuppose agency, but Korsgaard does not claim that membership of the noumenal world is not the same as being an agent. Agency, again, is autonomous efficacy, whereas said membership is a deliberative standpoint, the possession of which need not necessarily turn a person into an agent.

¹⁹⁶ KE 176.

¹⁹⁷ KE 178. Cf. MM6:395.

¹⁹⁸ G4:428.

¹⁹⁹ MM6:429.

²⁰⁰ See, for example, MM6:394.

²⁰¹ See especially KE 182-183.

and to overcome inclination. Even if the argument is correct, then, Enoch's shmagent could still ask why he should care about freedom and agency. Couldn't he simply content himself with faux freedom?

As it stands, the Korsgaardian answer outlined above seems to lack the means to provide a satisfactory answer to the shmagent's question. And it has yet another problem, which in fact seems to be the underlying problem of the issues above. Korsgaard, as I have tried to show, argues that we need a reason to be agents and that this reason has to do with morality's higher, legislative status. The issue is that Korsgaard takes a few steps too many. In the earlier bits of the argument, she has already shown that our rational nature gives us access to the noumenal, legislative and moral standpoint. Consequently, because this seems to entail that already are part of the noumenal world insofar as we do anything rational, it is redundant to argue that we *should* be part of the noumenal world. Furthermore, insofar as we already have access to the noumenal domain and insofar as that domain intrinsically places us under certain obligations, there is no need to posit morality's supposed higher status as a further reason to be moral. We would then already be under an obligation to engage in agential, moral behavior, even though we often might feel our inclinations pulling us in another direction.

A second issue is that, at least according to Kant, morality is binding because it stems from our own will and rational nature.²⁰² I will explore this idea in more depth in the next section, but here suffice it to say that if Kant's assertion is correct, there is simply no need to ground morality and its normative force any further by assigning a higher and therefore supposedly desirable status to it.

The underlying problem with the potential Korsgaardian answer to Enoch's shmagency objection, I believe, is thus that it makes agency contingent on reasons which one may or may not accept.²⁰³ This strategy invites further shmagency objections, because it is perfectly reasonable to ask why one should commit oneself to such things as 'higher vocations' or 'moral freedom'. If morality and agency depend on such further ends, then whoever does not care about those further ends can safely ignore morality's and agency's call. In the next section, I will therefore attempt to develop Korsgaard's answer into what I will argue to be a more promising direction.

2.3 Anthropological necessity

Korsgaard's constitutivist theory has more cards up its sleeve than the ones that have been the topic of discussion thus far. Seeing as how the strategy of making agency reliant on reasons does not seem to lead to satisfactory results, I will use this section to take a different Korsgaardian route, using the basic elements and analyses from the first chapter. As a background, Korsgaard states that the necessity of agency is neither logical, nor rational, nor causal.²⁰⁴ According to Enoch, this leaves

²⁰² See, for example, MM6:417-418 and G4:460-461.

²⁰³ To reiterate, I have been discussing a construction of one of her essays, which was not directly written as a response to Enoch's objections. This is only one way Korsgaardian constitutivists could respond to his charges, and as I have tried to show, it is not the most promising and therefore also not the most desirable or likely line of thinking.

²⁰⁴ SC 1-2.

Korsgaard arguing for a rather vague and supposedly unconvincing type of necessity.²⁰⁵ In this section, however, I will put forth a type of necessity that I will argue matches the overall project of Korsgaard's constitutivism and can be read into her work – a type of necessity I will call *anthropological necessity*. The strategy I will employ with regards to Korsgaard's work thus mirrors her own strategy with regards to Kant's work: it is an attempt to interpret her work – this was the goal of the first chapter in particular – and to position this interpretation in a current debate, which is the goal of the current and next two sections.²⁰⁶

Taking a bit of an oxymoron as a starting point, the necessity cannot be too strict. As Onora O'Neill notes, nothing can force you into being autonomous in thought or action.²⁰⁷ The necessity of agency must be such that it is defeasible and non-invasive.²⁰⁸ Whatever necessity agency has, it cannot be so strong that people are necessarily autonomous and efficacious or that they are necessarily forced to try to be so. Hence, I propose that the Korsgaardian answer must be sought in what we already are: (imperfectly) rational beings.²⁰⁹ This nature, I will argue, puts us in a position in which we can turn ourselves into agents and build an identity of our own, without claiming that failing to do so is unavoidable.²¹⁰ As a starting point, again, human beings are already endowed with rational capacities. As seen before, this set of capacities for Korsgaard is not some mysterious property, but starts from the idea that we must act for reasons and that we must regard ourselves as potentially being in control and free.²¹¹ It also entails, for reasons discussed before, that human beings must set themselves certain ends. When confronted with a desire, we must decide whether or not, and for what reason, we want to satisfy it. This is the human condition, on Korsgaard's analysis, and since there is no escaping our nature, engaging in reflection about what we have reason to do is something we necessarily have to do.²¹² For that reason, I call the necessity of this sort 'anthropological necessity'. This necessity is not

²⁰⁵ Enoch, "Agency," 188f-189f.

²⁰⁶ As before, I take this analysis of Korsgaard's work being both interpretative and argumentative from Sem de Maagt's "Constructing Morality," 260f.

²⁰⁷ Onora O'Neill, "Autonomy, Plurality and Public Reason," in *New Essays on the History of Autonomy: A Collection Honoring J. B. Schneewind*, ed. Natalie Brender and Larry Krasnoff (Cambridge: Cambridge University Press, 2004), 181-194, 190. Cf. MM6:381-382.

²⁰⁸ By 'non-invasive' I mean that the necessity of agency must be such that it is not forced upon people in an undesirable way. That is to say, agency and its necessity should stem from features that people already have and that can be shown to somehow be desirable. By contrast and for example, if agency were such that it forced people into regarding themselves as free and regarding oneself as free would be undesirable and/or unnatural, then that would mean that agency is invasive.

²⁰⁹ I thus propose to interpret Korsgaard in a common constructivist or constitutivist way, which starts from the assumption that substantive moral norms can be justified by means of anthropology (cf. De Maagt, "Constructing Morality," 260-267).

²¹⁰ I will argue that Korsgaardian constitutivism should move from (a) this first claim that human reflective nature requires that they deliberate at least at a basic level to (b) the idea that successful deliberation requires the reflective endorsement process described in chapter 1 and to (c) the idea that successful deliberation thus is in line with the demands of the moral law, as also described in chapter 1.

²¹¹ Korsgaard does not devote much attention to the fact that people often seem to do things unreflectively. Her point is that the necessity of agency is not something that works *on* people, as if it were something that had a causal influence on them. Anthropological necessity, to use my own term for it, is a situation people are necessarily faced with, even though they may sometimes or even often fail to recognize it (cf. SC 1).

²¹² There are many ways to fail to reflect, which is why I have referred to it as 'loose necessity'. These different possibilities to fail to engage in reflection will feature later on in this section.

a *reason* for agency. Because it is part of human nature, it is not something one can agree or disagree with.²¹³ And it is not invasive, because anthropological necessity stems from what we already are. We already are human beings; therefore we already have rational capacities and therefore we already are capable of understanding ourselves as members of the intelligible, noumenal world.²¹⁴ Because I have been making these Korsgaardian claims before, suffice it to say that from the fact that we are rational beings we can argue to the conclusion that morality is binding for us. For reasons that have also been laid out before, the rational side of our nature requires us to regard ourselves as free and thus to accept the CI in order to maintain and develop that freedom.

A different way of putting the point of anthropological necessity is that we, as human beings, have the idea of morality: it is, in Kant's terms, a 'fact of reason' that we are conscious of morality.²¹⁵ Of course, this does not mean that human beings, empirically speaking, are necessarily conscious of morality – not in the general meaning and certainly not in the more technical, Kantian sense. What it does mean is that human beings are endowed with rational capacities and once their reason starts working, they necessarily become conscious of morality. That is, consciousness of morality and its demands is inseparably linked with rational activity.²¹⁶ Since it is natural for rational beings to think, reflect and employ reason, anthropological necessity entails that it is quite natural to think morally and thus to regard oneself as free and behave as an agent would. Hence, agency is not simply up to one's own discretion – it springs from rational capacities human beings naturally have. To put the point stronger, the 'fact of reason' entails that every normally functioning human being potentially has a conscience or the idea that he or she can be held morally accountable.²¹⁷

To summarize, anthropological necessity is the necessity for agency that constitutivism needs because it is non-invasive and in fact perfectly avoidable in practice.²¹⁸ This last addition is important because no serious constitutivist would commit herself to the claim that agency is so strictly necessary that the demands of morality are unavoidable, for the simple reason that it belongs to the concept of moral imperatives that they are avoidable. Thus, rational human nature is such that we have the idea of morality and agency and thus we are subject to moral requirements, even though we often disregard these duties. It is, of course, also possible to fail to understand that you in fact or potentially are an

²¹³ One can, of course, agree or disagree with it as an analysis of human nature, but that is a different matter.

²¹⁴ Cf. G4:453-454, 460 and 461.

²¹⁵ C2 5:29-30.

²¹⁶ Cf. Pauline Kleingeld, "Moral Consciousness and the 'Fact of Reason,'" in *Kant's Critique of Practical Reason: A Critical Guide*, ed. Andrews Reath and Jens Timmermann (Cambridge University Press, 2010), 55-72.

²¹⁷ For an introductory read on the 'fact of reason', see Michael Rohlf, "Immanuel Kant," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2020, 5.3. See also Kleingeld, "Moral Consciousness" and Jens Timmermann, "Reversal or Retreat? Kant's Deductions of Freedom and Morality," in *Kant's Critique of Practical Reason: A Critical Guide*, ed. Andrews Reath and Jens Timmermann (Cambridge University Press, 2010), 73-89.

²¹⁸ This might appear to be using the same contradiction I have pointed out in section 2.2, where I wrote (footnote 178) that it is contradictory to try to combine the claim that agency is necessary with the claim that one would need a reason to be(come) an agent. The difference with that previous Korsgaardian answer, however, is that anthropological necessity claims only the former: agency is necessary, with the caveat that agency, as discussed in chapter 1, is the outcome of successful deliberation, and deliberation (not successful deliberation) is necessary for normal human beings. When one fails to deliberate completely, then Korsgaard, on my anthropological necessity interpretation, must concede that one does not exist as an agent (SC 69-70).

agent.²¹⁹ In spite of the fact that agency cannot go beyond this defeasible status, its necessity is stronger than Enoch seems to recognize. That is, because we *can* regard ourselves as agents – we are rational beings and therefore must operate under the idea of freedom – we have moral duties, and these hold even though we are only imperfectly rational.²²⁰ As discussed in chapter 1, morality and its duties according to Korsgaard revolve around the cultivation of our agency: autonomy, and therefore *a fortiori* also morality, is constitutive of action and by acting we constitute ourselves as agents. The necessity of operating under the idea of freedom means that doing so is imperative or inescapable to all beings capable of doing so.

Taking the point further, Kant and Korsgaard seem to agree that agency is not only necessary; people might even have a *duty* to be or become agents.²²¹ Their arguments differ, but the gist of it is this: there are certain ends which are objective or which it is a duty to have, such as the cultivation of one's own humanity (i.e. rational capacities) or one's ability to set oneself ends. Because we supposedly have a duty to improve our ability to set ourselves ends, we also have a duty to try to be autonomous and efficacious, because those are closely related to the ability of setting oneself ends, as seen before. To repeat the point only briefly, the idea is that our nature is such that we can only act on the basis of reasons and that we must set ourselves ends, and that doing so requires that we operate under the idea of freedom and thus regard ourselves as autonomous and therefore also efficacious. Korsgaard's point is thus that our nature puts us in a position in which agency, in the end, is not only necessary, but also something which it is a duty to either obtain or cultivate. There is therefore a way in which we not only necessarily *are* agents, but also *ought* to be agents.²²²

Enoch here could reply that this is only part of the story of human nature. From Korsgaard's AS it already followed that human beings would accept the CI as their principle if they had been fully rational beings. But we are not. We also have a 'passive' side, subject to temptations, desires, inclinations et cetera.²²³ So even granting that anthropological necessity establishes that we (potentially) have an agential side, why should we give that side precedence? Kant seems to anticipate this question, writing that

since sensible inclinations of human beings tempt them to ends (the matter of choice) that can be contrary to duty, lawgiving reason can in turn check their influence only by a moral end set

²¹⁹ This point is inspired by Deryck Beyleveld and Marcus Düwell's *The Sole Fact of Pure Reason: Kant's Quasi-Ontological Argument for the Categorical Imperative* (De Gruyter, in print).

²²⁰ See, for instance, G4:453-454.

²²¹ MM6:385-387; KE 176-183. One way of understanding the existence of this duty is to see it in light of Korsgaard's (and, perhaps, Kant's) claim about membership of the noumenal world, which might ground the duty to be agents even if agency is unavoidable (or even if one fails to be an agent, since Korsgaard seems to leave open the possibility that 'members of the noumenal world' are not agents).

²²² As I will argue next, human beings can fail to engage in reflection and thus to constitute themselves as agents and this makes the duty to be agents important. However, it is doubtful whether emphasizing one's duty to be or become an agent is much more than preaching to the choir, which is why I will also discuss other ways of extending or strengthening the necessity of agency.

²²³ Cf. KE 177-178.

up against the ends of inclination, an end that must therefore be given *a priori*, independently of inclinations.²²⁴

This, of course, leaves the door open for further Enochian questions, such as why one should set oneself the moral end that opposes the ends of inclination. As I have already tried to argue, giving reasons to be moral will not work. A different strategy would be to argue that, insofar as we understand that we are agents, we cannot without contradiction choose to give in to ends contrary to duty.²²⁵ Alternatively, in a simpler form of argument, one could argue that one would need agential capacities to sacrifice one's agency.²²⁶ Thus, to choose the ends contrary to duty one would need rational, agential capacities and to employ these in order to do something that undermines agency constitutes either a contradiction or a lack of understanding what it means to be an agent.²²⁷ Thus if the shmagent willfully chooses to not be an agent, he must employ his rational capacities, reflect on what he has reason to do and then through the whole process described in this section (and in greater detail in the first chapter) he must eventually conclude that he must regard himself as a free, autonomous being and thus that he cannot be a shmagent. To clarify, Ferrero and possibly Beyleveld would argue that choosing to be a shmagent is somehow contradictory. If my anthropological necessity interpretation of her work is correct, I believe Korsgaard would have to disagree. The whole process of establishing what one has reason to do or be is part of human nature and the *outcome* of the process of reflective endorsement is autonomy and efficacy. These two things are constitutive of agency and thus agency is not the reflective starting point, but rather the unavoidable outcome of successful reflective endorsement. By reflectively establishing what one has reason to do, one makes oneself autonomous and efficacious and thus constitutes oneself as an agent. It is along these lines that anthropological necessity avoids making a contradictory move. Although at first sight Korsgaard seems to advocate the seemingly contradictory combination of assumptions that (a) agency is necessary and (b) agency depends on our actions, if my anthropological necessity interpretation is correct, she in fact proposes that action is necessary; that action should be understood as the result of successful reflective endorsement; and that agency thus is not necessary on its own.²²⁸

But what if Enoch should be read as stating that a true shmagent does not care about reflective endorsement or morality and simply intends to satisfy his desires? Or as stating that a true shmagent does not engage in reflective processes in any sufficient depth? In those cases, anthropological necessity indeed seems to have no grip on him and thus it is unable to turn someone into an agent rather than a shmagent. I will discuss these two options in the next two sections.

²²⁴ MM6:380-381. Italics in original.

²²⁵ This is roughly the strategy Deryck Beyleveld employs in "Transcendental Arguments".

²²⁶ I will discuss this strategy in section 2.4.

²²⁷ More nuanced versions of this argument can be found in Beyleveld and Düwell, *The Sole Fact of Pure Reason* and Luca Ferrero, "Constitutivism", for example. Their arguments, again, are more subtle than the crude version I have described, and will feature more in-depth in the following sections.

²²⁸ I thank Micha Werner for pointing out the potential contradiction here. I will return to it in 2.5.

The hypothetical status of desires

Anthropological necessity establishes the necessity of agency by arguing that it results from our natural, rational capacities. But, again, human beings also have a non-rational side, which includes, among other things, inclinations and the faculty of desire. And it seems quite possible to live one's life by listening mostly or even exclusively to this side of our nature. Enoch's point of course is that a shmagent is a lot like an agent, and since someone living only to fulfill his desires is nothing like an agent, the person in question would, strictly speaking, not qualify as a shmagent. But on the assumption that Enoch's concern is with the necessity of agency, I take it that the possibility of living a life that centers around gratification would count in favor of his general project.

There are two distinct ways of understanding how one could live such a life. The first is to say that someone chasing the satisfaction of his desires is what could be called a rational egoist.²²⁹ On this interpretation, the 'shmagent' operates primarily on the aim of maximizing the satisfaction of his desires. Assuming that he is in principle a normal human being and thus has rational capacities, he generally ignores his rational capacities, focusing instead on the sole end of satisfying his desires. He thus cares about the interests of other people mostly insofar as it benefits himself to do so and usually ascribes only instrumental value to others. Because he does not care about his agential capacities, it would be pointless to argue that he in fact makes self-love his maxim and that he cannot do so without sacrificing his freedom. But there is a different problem with rational egoism. Consider the following thought experiment, which Korsgaard attributes to Rawls:

A man is going away to fight in a war, in which he may possibly die. The night before he leaves, the devil comes and offers him a choice. Either while he is away, his family will thrive and flourish, but he will get word that they are suffering and miserable; or while he is away his family will suffer and be miserable, but he will get word they are thriving and happy. He must choose now, and of course he will be made to forget that his conversation with the devil and the choice it resulted in ever took place.²³⁰

On the assumption that he really does care about his family, the shmagent as rational egoist is thus confronted with a choice: he can either let his family thrive while he believes that they do not, or he can let his family not thrive even though he will believe that they do. Subjectively, the latter choice would further the rational egoist's goal, because this would make him happier. But because even as a rational egoist he may still feel affection for his family and thus he might also choose the former.²³¹ The problem therefore is that he must decide what matters most to him. He needs to establish what weighs heavier here, his love for his family or the satisfaction of his desires. And according to the transcendental argument from the first chapter, making such a choice about what matters to you

²²⁹ See, for example, SC 166-168.

²³⁰ SC 167-168.

²³¹ Even a strict rational egoist might love his family, either (i) because he (as said) must be assumed to be a reasonably well-functioning human being and should thus be assumed to be capable of having feelings of love, or (ii) simply because he believes loving his family will help him satisfy his desires in the long run.

involves an appeal to your practical identities and, for example, an attempt to settle which matters most in this case. The point is this: if in life we are confronted with situations in which certain things that matter to us are in conflict with the satisfaction of our desires, as is not seldom the case, then we need to establish what we have reason to give precedence. And that requires self-scrutiny, reflective endorsement and other such things through which one constitutes oneself as an agent. This means that the shmagent as rational egoist must engage in reflection, thus pulling him firmly into agential territory. As for any human being, no desire has direct normative force and thus he must reflectively decide which desire he will try to satisfy and for what reason. Hence, interpreting shmagency as rational egoism is not a suitable way of circumventing my Korsgaardian answer to Enoch.

The second way of interpreting the non-rational type of shmagency is to regard it as what Korsgaard with Plato calls being a 'democratic person' or what Frankfurt in a similar vein calls a 'wanton'.²³² The 'democratic person', according to Korsgaard, listens only to what desire he happens to have on a given moment. Each desire is worth as much as the next and this means that the democratic person has no means at his disposal to organize his life.²³³ Suppose, for instance, that while writing this chapter, the sun comes out and I feel a desire to go outside. But before I put my shoes on, I realize that I would like a coffee, so I walk over to the coffee machine. And so my day unfolds and I will probably wind up getting very little writing done, not enjoying the sun, probably not enjoying a coffee, and so forth. It is of course possible that it all works out better and I *do* get enough work done, but the point is that this would be no more than an accident – if I truly act on whatever desire I have, I will have no way of grabbing myself by the scruff of the neck and making sure that I do what I have to do.²³⁴ As an aside, Korsgaard is willing to grant that such a person could still have friends with whom he could go out to drink a beer, but it seems doubtful whether the democratic person would have the right desires at the right time often enough to do what friendship requires. Simply put, it would be nearly impossible to be friends with a democratic person because such a person would not have a meaningful understanding of being your friend – they would only behave as your friend insofar as they happen to have a desire to do so.

What all this shows, however, is only that living life by following whatever desire you have is costly – not that it is impossible. But there is a real problem underlying both types of non-rational shmagency. Again taking my cue from Beyleveld and Düwell,²³⁵ the issue is this: if someone decides to try to satisfy his desires, he will act under the influence of the hypothetical imperative. This imperative represents the 'practical necessity of a possible action as a means to achieving something else that one

²³² SC 168-169 and Frankfurt, "Freedom," 11, respectively (it could also be the particularistic will O'Neill and Wood discuss, but that might be too much of a stretch (O'Neill, "Autonomy," 185-186 and Wood, *Kant's Ethical Thought*, 156)). Note that this interpretation of shmagency works both if one assumes that the shmagent is in some way 'determined' by his desires and if one assumes that the shmagent simply does not devote any attention to the normative status of his desires. That is to say, this interpretation is not necessarily committed to assuming determinism about desires.

²³³ Contrary to the rational egoist, who at least consciously decides and thus makes it his principle to follow desires.

²³⁴ I owe this line of thought to Korsgaard's SC 168-169.

²³⁵ Beyleveld and Düwell, *Sole Fact*.

wills', as Kant puts it.²³⁶ Düwell and Beyleveld argue that in order to recognize this principle you must be able to regard yourself as a member of the 'world of understanding', or the 'phenomenal world' as Korsgaard calls it. They advocate the view that if you see that M is a necessary means to E and you know that you desire E, then you ought to pursue M (or, perhaps, stop striving to achieve E). This entails, according to Beyleveld and Düwell, that you know that E is not something you necessarily strive for. They do not argue for this point in much detail, but it might be understood in light of Korsgaard's claims about the reflectiveness of human nature: when confronted with a desire, a human being's (self)conscious nature puts him at a certain distance from this desire, such that he knows that satisfaction of this desire is not immediately necessary. Desires cannot immediately steer the will, which people may come to recognize through their reflective nature, a claim Kant also makes. I will quote him at length, because he puts the point quite vividly:

Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him.²³⁷

No matter how strong the desire, then, human beings have access to (the idea of) morality and *a fortiori* can regard themselves as free. In the present context this means that it is in fact not true that a 'shmagent' – again, the types under discussion here do not meet Enoch's shmagency-criteria – can act only on the basis of desire and simply ignore his rational side. As soon as the shmagent becomes aware of the fact that he must deliberate about what he wants to do, he must regard himself as free and (potentially) in control of his desires. To counter the straightforward Enochian question here, this is not to say that he must care about substantive Kantian freedom or moral principles. Rather, it means that even human beings who live their lives chasing the satisfaction of their desires must engage in

²³⁶ See, for example, G4:414. On Korsgaard's interpretation, the hypothetical imperative helps one be or become the cause of the thing one wills (SC 72). Thus if one wills to write a paper and is then confronted with a desire to go do something else, the hypothetical imperative is binding because it dictates that one be the cause of the ends one wills and thus it prescribes that one continue working on the paper rather than letting oneself be distracted.

²³⁷ C2 5:30.

deliberation and that such deliberation, as argued before, implicitly involves such substantive Kantian principles.²³⁸

One issue with this line of thought is that someone might still be a shmagent as long as no one makes him aware of the contingency of his desires. That is, Kant invokes a second person who asks the lustful person what he would choose in the given situation, but what if no such person is around and the shmagent blissfully continues to chase his desires? While I do not believe it to be absolutely impossible, it seems sufficiently likely that one will occasionally have conflicting desires and that one will notice if such a conflict arises.²³⁹ And when confronted with conflicting desires or ones that will have undesirable consequences in the future, the shmagent will have to decide what to do. This places him in a position of freedom, where he can and must decide what matters most to him. He must either choose to make it his principle to simply follow the strongest desire, or he must choose between two conflicting desires. This latter option might be as basic as ‘I do not care enough about desire X to not follow desire Y’, but that would still be an instance of reflective thinking, which, if carried out properly,²⁴⁰ will lead to autonomy and agency. Interpreting the shmagency objection in the ways outlined in this section thus cannot turn it into a successful argument against anthropological necessity.

The importance of wholeheartedness

From a Kantian point of view, Enoch’s claims about agency and half-hearted agency may seem unproblematic. There are good reasons to believe that, according to Kant and numerous constitutivists, shmagency is impossible for reasons outlined above: in the end, you would need agential capacities even to try to satisfy your desires and so would need said capacities in order to be a shmagent. But in a Korsgaardian framework, agency is not as strong and perhaps not self-sufficient.²⁴¹ Enoch’s points about agency therefore do apply; if agency depends on one’s actions and engagement in reflective processes, then one may still ask why one should be(come) an agent. It is still an open question whether Korsgaard’s theory has the means to answer one last iteration of the shmagency objection: the half-hearted form of agency that Enoch also discusses.²⁴² In the previous section, I have tried to prove that shmagency as a deliberate choice by a rational being is impossible, and that it is impossible to evade agency by ignoring one’s rational side and living a life centered solely around the satisfaction of desire. In this current section I will try to show that shmagency also cannot be understood as a lack of deep reflection or serious participation in the ‘game’ of agency. I will try to

²³⁸ As before, one can reasonably ask if it is true that reflective endorsement requires Kantian moral principles, but the mere fact (or stipulation) that a shmagent does not or need not care about such principles does not refute my point here.

²³⁹ Even if one fails to notice, one should still be held accountable, Korsgaard seems to claim. See also footnote 253.

²⁴⁰ See also 2.5.

²⁴¹ Luca Ferrero, by contrast, advocates the view that agency can be self-sufficient. His arguments will be the topic of section 2.4.

²⁴² Enoch, “Shmagency,” 210-212, for example.

show that it is impossible to live life without sincerely, wholeheartedly trying to taking control.²⁴³ Before doing so, however, I will first go into shmagency again, because there is a particular version of shmagency that I have not yet discussed and that is important to the point I intend to make.

Enoch, as said, proposes two types of shmagency: the strict, metanormative one I have discussed at the beginning of this chapter and shmagency at the normative level. Enoch likens this last type to playing chess without having a reason to play the game.²⁴⁴ The point, it seems to me, might be read as a counterargument to the Kantian line of thinking, and it revolves around the idea that even if there are constitutive standards of certain activities, they do not suffice to ground our commitment to those activities. So even if there are constitutive standards of agency, it is still necessary to show why we should be engaged in agency; the constitutivist, according to Enoch, needs to provide a further, non-constitutivist reason for being an agent, because constitutivism itself cannot show the necessity of (adhering to the constitutive principles of) agency.²⁴⁵ By way of example he discusses chess: suppose you are playing a game of chess, but you do not care about the game or who wins. Enoch thinks this would mean that you have no real reason to try to checkmate your opponent. He argues that even when a metanormative theorist were to come along and explain to you that checkmating your opponent is a constitutive rule of chess, you would not have a reason to do so. Because you do not really care about the game you are playing, it does not matter to you whether it can really be classified as chess. The same goes for agency, Enoch states, and thus constitutivism at the very least cannot tell the whole story: if constitutivism needs a further, external reason for agency before agency's constitutive principles can normatively kick in, then constitutivism is only part of the story and we are back at square one.²⁴⁶

There is an important addition to be made to Enoch's point. Continuing with the chess-example, Enoch argues that there could be all sorts of games of not-quite-chess we could be playing and which we cannot clearly distinguish between. So while moving chess pieces across the board, it is not clear whether one is playing chess, chess*, chess**, et cetera. Enoch thinks this shows that 'engaging in some activity – satisfying some relevant descriptive criteria – [does not suffice] for having reason to direct oneself at its constitutive aim'.²⁴⁷ He continues by stating that constitutivism is in fact more nuanced than his chess example suggests, because constitutivism is the view that playing chess already means caring about the fact that one is playing chess.²⁴⁸ He rightly attributes to constitutivism the idea that one only engages in actions or chess (rather than something that could be chess, chess*, and so on) if one already cares about it being an action. Hence he concedes the point that you would not even count as playing chess 'unless you are committed to achieving [the constitutive aim of chess], unless you care about checkmating your opponent'.²⁴⁹ Yet now we have arrived at something 'terribly

²⁴³ This, based on the Korsgaardian theory I have been developing here, should be the general attitude, which of course means that one can fail to do so on more than one occasion without harming the general need to sincerely reflect. See also footnote 258.

²⁴⁴ Enoch, "Agency," 185-187.

²⁴⁵ *Ibid.*, 186.

²⁴⁶ The point put forth in these paragraphs is taken from *ibid.*, 185-187 and "Shmagency," 209-212.

²⁴⁷ Enoch, "Shmagency," 211.

²⁴⁸ *Ibid.*, 212.

²⁴⁹ *Ibid.*

uninteresting', Enoch writes, because constitutivism concerns itself with asking whether someone playing something a lot like chess *really* is playing chess, or whether someone who is a lot like an agent *really* counts as an agent.²⁵⁰ This ultimately leads him to criticize constitutivists who claim that real action is nothing like 'shmaction' or that real agency is nothing like shmagency. Such claims, Enoch argues, are 'too much to swallow',²⁵¹ meaning that Enoch advocates the view that action is a subtype of behavior and that there can be other types of behavior that are close enough to action, even if they do not quite make the cut. Thus, Enoch seems to think that it is somewhat silly and rather uninteresting to ask what truly qualifies as action or agency.

Enoch's claim that constitutivism ends up asking uninteresting questions signals a misunderstanding of constitutivism's project.²⁵² Of course, it is not interesting to ask whether someone who is half-heartedly engaging in something that looks like chess is indeed playing chess. But whether or not someone is really engaged in agency and action *is* interesting and of great importance. As discussed before, the domain of agency and action is the domain of morality and autonomy. To restate the point, Enoch argues that a lot of our behavior might not quite be action in the technical sense, but that there is no problem here.²⁵³ Such behavior is good enough as it is and there is no point in asking whether it lives up to the constitutive aims and principles of action. Enoch thus finds that constitutivism places a lot (too much, rather) of emphasis on the difference between behavior that is action in the technical sense and behavior that is a lot *like* action. Constitutivist theorists in effect try to place a boundary between these two things. They would be committed to the view that action is a special subtype of behavior and that everything that tries to imitate that without adhering to its constitutive principles – autonomy and efficacy – is 'shmaction'. Again, in a last step, Enoch claims that this difference is 'uninteresting'. Where the difference between wholeheartedly playing chess and playing chess with one's child is the topic, it is easy to see why the difference would not be terribly interesting. Because Enoch grants constitutivism its claims about the constitutive principles of action,²⁵⁴ one would expect him to know and accept that this means that morality and action are inseparably combined and then his claim that it is boring to ask whether something truly qualifies as action is rather surprising. His questions with regard to half-hearted agency may still stand, however, and they will be the topic of the remainder of this section.

When anything more important than chess is at stake, the constitutivist project of finding out whether or not something adheres to that activity's constitutive principles, again, *is* important. An example might help shed light on the issue. For the sake of clarity, I will stick with Enoch's chess

²⁵⁰ Ibid., 214.

²⁵¹ Ibid., 215.

²⁵² Enoch, again, seems to be interpreting agency's necessity as a conceptual necessity, meaning that agency may indeed have constitutive standards and that one can only be a real agent insofar as one meets those standards. Enoch's objection then is that it does not matter to a shmagent whether or not he meets those standards, because he is indifferent to (his own or other people's) judgments about whether or not he deserves to be called an agent (I thank Micha Werner for pointing me towards this 'conceptual necessity' interpretation). As I will try to argue in the remainder of this section, it does matter whether one meets the standards of agency and thus whether or not one is an agent.

²⁵³ See, for example, "Shmagency", 214-215.

²⁵⁴ Ibid., 214.

example, but add a different layer. The person in Enoch's example is playing chess half-heartedly: he knows the constitutive rules of chess, but does not care enough about the game he is playing to accept its rules as normative, and thus may pass up the chance to checkmate his child when that opportunity presents itself. I see three ways of understanding this half-heartedness. First, he may not care about chess enough to try to win, because he wants his child to have an enjoyable game. Second, he may not care about the rules of chess because he has decided (consciously) to be half-hearted about reflection and most or all other things in life. Third, he may not care about the rules of chess because he has always (unconsciously) been half-hearted about reflection and most or all other things in life.

In the first scenario, the parent is simply playing along. He could perhaps easily checkmate his 'opponent' but is not committed to this goal because he is not interested in chess's constitutive principles. He has different commitments (being a good parent) that outweigh his commitment to playing chess exactly the way its constitutive principles prescribe. By letting the reasons given by one's being a parent outweigh the reasons one's playing chess give, however, one constitutes oneself as a parent and, arguably, also as an agent. Insofar as half-hearted commitment to an activity is caused by a wholehearted (and reflectively endorsed) commitment to something else, then, half-heartedness is not an interpretation of shmagency that will turn it into an effective objection to constitutivism.

Alternatively (in the second scenario), the parent in question might not care about the constitutive rules of chess because he has decided not to care a lot about most things in life, or not to overthink things. To clarify, Korsgaard does not claim that every decision should follow only after long and thorough reflection,²⁵⁵ but there is a large gap between knowing when to stop engaging in reflective endorsement and deciding (beforehand) to reflect only superficially, as 'not overthinking it' suggests. Based on what I have been saying thus far, the evident question here is how one would go about making it one's principle not to reflect seriously. There are two ways of understanding why making such a thing one's principle is not possible. First, one can only decide what principles one can accept by exposing them to the reflective endorsement test. One would then have to establish whether one could will the corresponding maxim – 'I will not reflect seriously, because not doing so serves my interests', for example – as a universal law. To name only two problems briefly here, such a maxim would lead (a) to a teleological contradiction because one would use one's reflective nature to turn oneself into an unreflective being; or (b) to a practical contradiction, because when universalized, the maxim of not reflecting seriously would stop one's need for maxims such as the one in question. Once one stops reflecting, one also stops needing maxims, and thus maxim is self-destructive. The reflective process is such that it cannot be used to reflectively decide never to be wholehearted about reflection.

Second, there is a further Korsgaardian, transcendental argument that counters both the second and third scenario. No matter whether one consciously or unconsciously comes to never deeply reflect or wholeheartedly care about one's activities, the human predicament is such that people must form an identity of their own.²⁵⁶ Forming an identity requires that one makes an effort to be the cause of one's ends – it requires that one performs actions. Actions require reasons, which in turn requires

²⁵⁵ See, for example, SC 126.

²⁵⁶ Since this is an analysis of human nature, Korsgaard's point seems to be that this holds true no matter if one accepts this or tries to remain passive about it.

reflection. Therefore, reflection is necessary for building an identity. One way of avoiding the force of this transcendental argument, of course, is the deny of the necessity of the starting point, essentially countering Korsgaard's analysis of human nature. One would then have to show that it is possible to live a (sufficiently normal) human life without having a personal identity. At first sight, this might seem possible. Without a personal identity and without distinct qualities, one might be rather lonely, because there is no clear reason to start a relationship or friendship with such a person, but it would not be impossible. From the first-person perspective, however, life without an identity or without caring is impossible. One need only consider the fact that in order to live, human beings need food. Yet from the transcendental argument I have discussed in 1.1, it follows that food is not valuable in itself and that one can only value one's food insofar as one, ultimately, values oneself. It is impossible to continue to live without ascribing a very basic level of value to oneself.

A different, more extreme, example may further show the need for wholeheartedness and valuing. In the Netherlands, euthanasia is legal if several strict criteria are met, one of which is that the person in question has sincerely judged that he would prefer to end his life. The point is simple: it matters a great deal whether one has seriously reflected on it and whether truly believes that euthanasia is the best decision given one's circumstances. The extent to which one succeeds in doing so is a matter of degree, but the point remains that one can only successfully apply by convincing the relevant parties of having seriously thought about it, or by deceiving them into believing so and avoiding any suspicion of deception. What this is meant to show is that, as with the previous examples, it matters that one is genuinely committed, because otherwise one cannot expect to achieve one's goals or build an identity.

To return to shmagency, the examples show that shmagency interpreted as half-hearted agency is not a viable option: when something of some importance is at stake, it matters that someone is serious about his commitment to that activity and therefore also (at least implicitly) about adhering to its constitutive standards. To clarify, the core idea is that half-heartedness is not a strategy or attitude that will help one achieve one's goals, and thus not that half-heartedness is immoral.²⁵⁷ For practical reasons, one must set oneself certain ends, and once those ends are in place, one must also will the means and one must also want to be the cause of one's ends. Of course one can and will fail on many occasions, but if my understanding of Korsgaard is correct, the core idea is that one must be held accountable if one does not at least make an effort to be the cause of one's ends. To put it in Korsgaard's terms, it is one's job to make oneself into an agent, to give oneself an identity and to become a person, and, in short, to 'pick up the reins and take control'.²⁵⁸

Furthermore, there is reason to believe that Enoch's half-hearted chess player in fact is not a half-hearted chess player, but a devoted parent, adhering nicely to the constitutive principles of

²⁵⁷ Cf. SC 180.

²⁵⁸ SC 175. According to Korsgaard, we should not 'pick up the reins' because we must try to do the right thing, but because we, as human beings, must act. Action, next, requires that we make ourselves the cause of our ends and that, in turn, requires that we deliberate and establish what matters most to us given the circumstances. And by doing so, we unify our wills, and this unity of the will is the reason, according to Korsgaard, why people must take control – unifying oneself is the only way to successfully confronting the challenges of the human condition of having to act (see, for example, SC 26).

parenthood. Enoch's claim that it does not matter whether someone really counts as playing chess is false and diverts attention from the fact that his example may portray a genuine case of commitment to constitutive principles, even though they are not the constitutive principles of chess. It matters a great deal whether or not one follows the constitutive principles of a certain action and what one's commitment to those principles is, and Enoch's example of the supposedly half-hearted chess player fails to show that half-heartedness is even a possible way of being a shmagent.

To conclude this section, I will briefly discuss a different line of thought in Korsgaard's work, which goes like this. As we have seen before, practical identities are key to the process of deciding what one has reason to do. So when deciding what type of chess one has reason to play (competitive or low-key), one must see what one's practical identity of being a parent, for example, gives one most reason to do. Playing a low-key game of chess in which someone lets his child win shows a good understanding of what being a parent gives him reason to do. Furthermore, it shows that he values his parenthood more than his win to loss ratio as a chess player and thus that he has his priorities straight. He knows what identity he must value over other practical identities in these circumstances and thus he knows how to prioritize between identities and the reasons they give him. I will return to this in more depth later, but for now suffice it to say that this means that he is good at being a person and that according to Korsgaard, this makes him a good person.²⁵⁹ In short, Enoch's half-hearted chess player is in fact a parent with a healthy understanding of what he has most reason to do and thus Enoch has not proven that half-heartedness is a workable alternative to agency.

The underdevelopment of shmagency

I will end this section on anthropological necessity as an explanation for the necessity of agency in Korsgaard's theory with a brief recap and with a more direct return to shmagency.²⁶⁰ To summarize the points I have been making above, shmagency as a deliberate *choice* is ruled out; human beings are rational by nature and it is impossible to utilize these capacities to become a shmagent. If carried out properly, the reflective process cannot but lead to autonomy and agency. The interpretation for the necessity of agency I have been exploring has been this: human beings are reflective rational beings by nature and this means that they, insofar as they employ their rational capacities, must regard themselves as free and efficacious. This means that shmagency on Enoch's understanding is impossible – anthropological necessity rules out the possibility that we lack the constitutive principles of agency and still be otherwise very similar to agents. Agency is not entirely up to one's own discretion, because the preconditions come naturally and cannot simply be ignored without leading to types of behavior with which one cannot constitute oneself and which are therefore unlike the type of behavior agents display: action.

Successful shmagency therefore should not be sought on the rational side of our nature. I have, however, argued that it also cannot be found on the non-rational side. If shmagency is understood as desire-satisfaction, then it will not work, because desire-satisfaction also requires that one regards

²⁵⁹ SC 214. I will return to this in 3.3.

²⁶⁰ I will here discuss the supposed possibility of shmagency and at the end of 2.5 I will discuss the shmagency debate more generally.

oneself as free and that one be able to establish which desire is most important. And for reasons given in the previous section, shmagency as half-hearted agency also does not work. There will be plenty of slipups and a lot of half-heartedness over the course of our lives, but half-heartedness as a general attitude that is supposed to replace actual agency will not do. Hence there are no viable options for shmagency to get off the ground. Even though anthropological necessity establishes only a partial (rational) necessity of agency, this necessity also carries over to the non-rational side of our nature and thus there is no room for shmagency. To conclude, in a somewhat obscure section Enoch argues that showing that shmagency is impossible does not help constitutivism's own more positive aspiration of establishing the necessity of agency.²⁶¹ He is, of course, right. Refuting the shmagency objection is not an end in itself and does not help establish the necessity of agency. But in the process of arguing against the supposed possibility of shmagency, I have tried to also establish the necessity of agency.

As Enoch admits, his first definition of shmagency leaves something to be desired,²⁶² but there is a further and perhaps more important issue with shmagency in general. Enoch seems to consider shmagency to be something that can exist in isolation, as if it would enable one to disregard the constitutive principles of agency and still lead an otherwise completely normal life. This is problematic for several reasons. First, for reasons I have been discussing, human beings need to reflect, and if the Korsgaardian position I have been describing is correct, (successful) reflection leads to agency. It is doubtful whether there could be normal human beings who lack such reflective capacities. Second, it seems quite simply false that one could lack the constitutive principles of agency and still be similar to an agent.²⁶³ To make a simple constitutivist move: assuming that shmagency can be successfully defined and assuming that shmagency is possible, then we could ask what makes a shmagent into what he is? What is constitutive of shmagency? These are important yet unanswered questions. In order to show that shmagency truly is something like (but not identical to) agency and in order to show that shmagency is possible, they must be answered. Only then will shmagency become a meaningful and properly developed concept to discuss.

Furthermore, agency makes sense only in a larger conceptual network and within the practical domain, as Luca Ferrero points out.²⁶⁴ Shmagency, if it is to be sufficiently like agency, would need a similar network of concepts which would have to be explained in non-agential terms.²⁶⁵ As I have tried to show, action is possible only for agents, so shmagency would need an alternative of its own, something shmagents do: *shmaction*. And it would, arguably, require *shmautonomy*, allowing the shmagent to be not-quite-autonomous but still close enough to resemble agential autonomy. Shmagency thus cannot exist as an isolated and harmless alternative to agency. It needs much more theoretical scaffolding than Enoch seems to recognize.

²⁶¹ Enoch, "Shmagency," 220.

²⁶² Ibid., 214.

²⁶³ This is not to say that human beings must also *care* about the constitutive principles of agency, as I will argue in 2.5.

²⁶⁴ Ferrero, "Constitutivism," 321.

²⁶⁵ Cf. *ibid.*, 321.

In the practical domain, a similar point applies. In “Deciding to believe”, Bernard Williams describes the potential consequences of believing at will.²⁶⁶ Suppose a man’s son has most likely drowned at sea. Believing that to be the case, however, is too much to bear for the man. But if he decides to believe that his son is still alive, he will have to change a whole web of beliefs. When he is called by the ship’s captain, for instance, he will have to find a way not to trust the captain’s claim that his son fell of the ship. When the coastguard calls to inform him that they have found a man matching his son’s description, he will have to not believe them, and so on. Something similar would apply to shmagents. A shmagent, by Enoch’s definition, lacks the constitutive principles of agency and thus lacks autonomy.²⁶⁷ Returning to the example of the chess-playing parent, this means that a shmagent cannot fully value his parenthood as a practical identity he reflectively endorses, because reflective endorsement and autonomy are inseparably combined. He would have a rather contingent and – as I have called it earlier – half-hearted commitment to this parenthood. But if he somehow decides to do at least some of the things parents typically do nonetheless, then he will also have to deceive his child into believing that he really is committed; that he values the fact that he is a father; that he consciously endorses his being a father as a practical (reason-giving) identity, and so on. To put the point in Frankfurt-like terms, people may be expected to care that they care. Children, to stick with my example, may expect their parents to care about being parents, and such caring requires that they endorse their parenthood as a practical identity (which, in turn, requires reflection and leads to, ultimately, autonomy and agency). Shmagency cannot possibly exist as an isolated phenomenon that somehow, by stipulation, still bears close resemblance to agency. Both on the theoretical and practical level, shmagency has serious further implications that Enoch fails to take into consideration and this shows not only why shmagency, if it were possible, would be rather undesirable, but also why it would need to be supported by a lot more theoretical, conceptual work.

2.4 The indirectness-objection

In the previous sections I have tried to show that it is possible to combine different parts of Korsgaard’s work to form a theory about the necessity of agency. Along the way I have also mentioned a different theory about said necessity, which is more direct and perhaps stronger than the one I have proposed. This theory will be the topic of this short section. Its basic idea is that agency can account for its own necessity and/or that any break away from agency requires agential capacities. Following Enoch, one might appropriately call it *dialectical inescapability*.²⁶⁸ This approach, as I will try to show, is a lot more direct than anthropological necessity and thus seems to offer at least one advantage over the approach I have been describing. It can be found, for example, in David Velleman’s *How We Get Along*, and a

²⁶⁶ Bernard Williams, “Deciding to Believe,” in *Problems of the Self* (Cambridge: Cambridge University Press, 1973), 136-151, esp. 149-151.

²⁶⁷ Enoch, “Agency,” 179.

²⁶⁸ See Enoch, “Shmagency,” 215 and 218.

similar yet distinct version is advocated by Düwell and Beyleveld in *The Sole Fact of Pure Reason*.²⁶⁹ Velleman sums up the gist of the argument clearly:

To ask “why should I have the aim of making sense?” is to reveal that you already have it. If you don’t seek to do what makes sense, then you are not in the business of practical reasoning, and so you cannot demand reasons for acting or aiming.²⁷⁰

By ‘making sense’ here Velleman means self-understanding, which in turn is a constitutive principle of agency according to Velleman.²⁷¹ In essence, the point is thus that the question why one should set oneself one of agency’s constitutive aims and thus why one should ultimately be an agent can only be asked by means of practical reasoning and thus only by showing that you already are an agent.

Luca Ferrero takes a similar route, arguing as follows:

Because of the special status of agency – the inescapability of the enterprise of agency and the indispensability of the concept of agency – the question whether there is reason to be an agent cannot be raised and answered outside of agency.²⁷²

There are two important things to note here. First, Ferrero’s definition of agency differs from the one Korsgaard uses. He claims that agency is the capacity to base your behavior on certain reasons you recognize for action and the capacity to ‘engage in the practice of giving and asking for these reasons’,²⁷³ whereas for Korsgaard agency is efficacious autonomy. Second and relatedly, it is at least debatable whether Ferrero’s account of agency can be reconciled with Korsgaard’s theory of agency and therefore it is also debatable whether it could serve as an alternative to anthropological necessity in a Korsgaardian framework. For Korsgaard, agency is constituted by actions and these require prior reflection on what one has reason to do. Ferrero, on the other hand, places agency at the start or even sees it as identical to the reflective process, raising the question how or if his theory of agency can really be reconciled with Korsgaard’s constitutivism.

Nonetheless, with this definition in mind, Ferrero goes on to argue that agency has the ‘highest jurisdiction’, meaning that all other enterprises and activities are subject to agency.²⁷⁴ This means that practical reflection on whether one has reason to engage in a certain activity ultimately is an agential activity. Since this type of reflection seems crucial to the proper functioning of human beings, this

²⁶⁹ James David Velleman, *How We Get Along*. The argument Düwell and Beyleveld propose can also be found in Beyleveld’s “Transcendental Arguments”. I will not discuss this argument here, because it focuses on agential self-understanding rather than bare agency.

²⁷⁰ Velleman, *How We Get Along*, 137.

²⁷¹ Ibid.

²⁷² Ferrero, “Constitutivism,” 322-323.

²⁷³ Ibid., 307.

²⁷⁴ Ibid., 308.

function of agency is one way of understanding why agency, according to Ferrero, is an indispensable concept.²⁷⁵

Agency is not only indispensable as a concept, Ferrero continues, but also inescapable. Because all other activities fall under agency, they are all open to reflective assessment; one can always reconsider one's reasons for participating in a given activity. Not so for agency. Agency is 'closed under the operation of reflective rational assessment', Ferrero writes, meaning that it is impossible to 'put agency on hold' and try to determine whether one has reason to be an agent.²⁷⁶ It is certainly possible to reflect on agency, but every reflection on agency also requires agential reflective capacities and therefore agency is 'closed'.²⁷⁷

To sum up the idea, Ferrero first assumes that human beings need reasons to engage in activities. He then argues that agency is the capacity with which one can recognize such reasons and reflect on whether one has reason to do something. The combination of these two things then makes agency indispensable for human beings. And since reflection on agency also requires agential capacities, agency is not only indispensable but also inescapable.²⁷⁸ If this argument for the necessity of agency works, then there is a much more direct route to grounding agency than the Korsgaardian one I have proposed in section 2.3. Assuming for the sake of the argument that Ferrero's theory can be modified so that it matches Korsgaard's constitutivism, the next question here is whether it is a satisfactory approach with which one can counter the shmagency objection. There are several reasons to believe that it is not. In the next section, I will try to explain that Ferrero's dialectical-necessity approach does not work as an answer to Enoch's challenges, and that it would run into more directly normative problems even if it were to be a successful refutation of Enoch's objection.

The shmagency trilemma

One issue potentially haunting dialectical necessity is one that I have already discussed in relation to anthropological necessity: the fact that defeating skeptical challenges will only get one so far in the positive project of establishing the validity of one's own claims. Enoch makes the point by giving the example of someone skeptical about the philosophical value of writing papers rather than books.²⁷⁹ She decides to express her worries in the form of a skeptical argument which she puts forth in a paper. Non-paper-skeptics could then easily point out that, apparently, she is not all that troubled by the limitations of philosophical papers, because she uses that very medium to make her point. However, according to Enoch such a response misses the point, because it says nothing about the actual

²⁷⁵ Another way of understanding his indispensability claim is that agency is a crucial building block for constitutivism (see "Constitutivism", 319-322). This might paradoxically make agency's necessity contingent on the necessity of constitutivism, but Ferrero does not entertain that possibility. On the relation between human nature and the need for agency, see also his "The Simple Constitutivist Move," *Philosophical Explorations* 22, no. 2 (May 4, 2019): 146-162, 155-156.

²⁷⁶ Ferrero, "Constitutivism," 308 and 309; Ferrero, "Constitutivist Move," 156; Luca Ferrero, "Inescapability Revisited," *Manuscripta* 41, no. 4 (December 2018): 113-158, 131.

²⁷⁷ Ferrero, "Constitutivism," 309.

²⁷⁸ *Ibid.*, 309.

²⁷⁹ Enoch, "Shmagency," 219-220.

argument the paper-skeptic makes.²⁸⁰ Put this way, Enoch's question is essentially a reiteration of his earlier claim that the skeptic or shmagent is entitled to use agential 'weapons' against agency.²⁸¹ If I understand his claims correctly, the point is as follows. Anti-skeptical arguments are not unlike ad hominem arguments, in that they often only show that a skeptic cannot take a stable position from which to launch his skeptical charges.²⁸² The problem then is that this would only show that the skeptic herself is inconsistent, which Enoch regards as saying little or nothing about the actual contents and validity of the skeptic's arguments. Enoch thus claims that the shmagency challenge should be taken seriously even if it is impossible to stably be a shmagent or to find a stable position from which to argue that shmagency is possible.²⁸³

To make a slight detour, the Korsgaardian account of agency I have formulated is not susceptible to this critique, because that account is in principle willing to concede that there can be meaningful philosophical argument about the necessity of agency. Assuming that rational reflection is something agents do, it is perfectly acceptable to the Korsgaardian theory that an aspiring shmagent uses said reflection to wonder why he should be an agent. The Korsgaardian account then replies, as I have tried to show, that one could, without contradiction, wonder why one should be an agent, but that there are no alternatives.

Ferrero's account has no such theory to back it up – not least because its claims about the necessity of agency are much stronger. Ferrero could respond to Enoch's point by stating that in asking whether one could be a shmagent one already uses agential methods of reflection and reason-giving. It would then somehow be forbidden or impossible to turn agential reflection against itself. Assuming for the sake of the argument that that is correct, it would still lead to further problems. On Ferrero's approach, one would reply to the shmagent that he cannot be a shmagent because he uses agential methods and thus inevitably shows that he *does* care about agency.²⁸⁴ But if someone asks why he *should* care about X, it is not a satisfactory answer to say that he already *does* care.²⁸⁵ Both Ferrero's and Korsgaard's (or rather: my Korsgaardian) theory about agency need to provide an answer to this further challenge. In section 2.5 I will try to provide a Korsgaardian answer, but within the current context suffice it to say that Ferrero, as far as I am aware, has not provided an answer, nor do I see any evident way for him to do so.

Even if Ferrero were to clarify why shmagents are in fact agents, and caring ones at that, this would confront his theory with a trilemma. To see this, suppose Ferrero succeeds in establishing a strong necessity of agency, which rules out shmagency because that would only show that one in fact already cares about one's agency. The problem then is that Ferrero also claims that agency is not

²⁸⁰ See Enoch, "Shmagency," 219-220.

²⁸¹ Enoch, "Agency," 184.

²⁸² See *ibid.*, 219-220 and Enoch, "Agency," 184-185.

²⁸³ Again, I might be misunderstanding Enoch's somewhat unclearly phrased argument here and will gladly point the reader to the passages referred to in the previous two footnotes.

²⁸⁴ As I will discuss in 3.2, my interpretation of Korsgaard differs from Ferrero's theory at least in that it is not committed to the claim that one should care about agency itself. It also differs from Ferrero's work in that it does not claim that to reflect about the need to be an agent means to already be employing agential capacities.

²⁸⁵ Cf. Enoch, "Shmagency," 212-213 and 215-217.

‘ontologically necessary’: human beings are not necessarily rational agents.²⁸⁶ Hence it is possible for a human being to not be an agent.²⁸⁷ We could therefore justifiably ask once more why human beings should be agents. Asking this question of course could mean that I, Bram Galenkamp, am an agent. But that reply, as I have argued with Enoch, does not answer the question, because it only shows that I am an agent and not why human beings in general should or would be agents. So seeing as how Ferrero defines agency as the capacity to ask for reasons, where does this leave a human being who never actualizes this capacity, or who lacks it altogether, or who simply does not care about reasons and agency? Why should or would he be(come) an agent?

Ferrero has three equally unappealing options here. First, he could argue that this type of human being is in fact not possible. However, that would require him to give up the claim that agency is not ontologically necessary and thus he would have to hold that every human being is necessarily an agent. Combined with his argument that agency is inescapable, this would mean that every human being is always an agent and can escape her agential plight only through sleep or death.

The second option is to argue that constitutivism has non-rational, non-agential means of convincing those who do not care about agency. Because Ferrero argues that human beings are not necessarily agents nor that they necessarily care about agency – if they did, they would presumably already be agents – those advocating his theory of agency would have to find a way to ‘convert’ non-agents into agents if they are to avoid the first horn of the trilemma. Again, since non-agents do not care about agency and thus about rational reflection, they would have to be convinced by non-rational means, and this ‘conversion’ would of course also need some justification.

Alternatively, the third option would be to acknowledge that agency is not ontologically necessary and that therefore human beings are not necessarily agents, and that it is not a problem if someone is not an agent. The point here is that Ferrero’s account could concede that there might be non-agents and that these cannot be convinced rationally. And if they also cannot or should not be convinced by non-rational means (thus avoiding the second horn), then those adhering to Ferrero’s theory would have to explain that non-agents are beyond saving. If someone who does not care too much about rational reflection qualifies as a non-agent, then the group of non-agents could in fact be quite large. And then it is not easy to argue that this is an attractive bullet to bite.

2.5 Anthropological necessity and the shmagency trilemma

Anthropological necessity does not run into the trilemma haunting Ferrero’s theory, or so I will try to show in this last section. To summarize the point about anthropological necessity very briefly, agency is necessary because human beings are rational beings by nature and their rational capacities enable

²⁸⁶ Ferrero, “Constitutivism,” 309. Although Ferrero does not define what he means by being ‘ontologically necessary’, I take him to mean that it is (perhaps empirically) possible to not have or use agential capacities and still be an otherwise normal human being.

²⁸⁷ Whether it is possible for a human being to be a shmagent is a different issue, which need not be settled for this argument to work.

them to regard themselves as free and thus as autonomous and, ultimately, as agents.²⁸⁸ As I have tried to show, the reflective process that links rational nature to agency is such that it cannot but lead to agency if it is properly carried out. The possibility of shmagency should therefore be sought in (what appears to be) non-reflection, such as the satisfaction of one's desires.²⁸⁹ This is a rather generous understanding of Enoch's shmagency claims, because he, as I have discussed, believes that shmagency is possible even under the stipulation that shmagency is very similar to agency. I have tried to show that shmagency is not possible if it must be very similar to agency, but that one might still interpret shmagency in a broader sense, such as half-heartedness or avoiding agency by focusing only on the satisfaction of one's desires. Even under the broader understanding, however, I have argued that shmagency is still impossible by trying to show that even the satisfaction of desires requires agential capacities and properties, meaning that even desire-satisfaction requires that one reflectively decide what one has reason to do.²⁹⁰ Assuming for the time being that that argument works, it remains an open question whether my anthropological necessity theory fares any better than Ferrero's theory at avoiding the shmagency trilemma.

Anthropological necessity can circumvent the shmagency-trilemma as follows. It is not threatened by the first horn, because it relies on what I have called a 'non-invasive' type of necessity. Anthropological necessity starts from the idea that human beings, qua rational beings, must regard themselves as free and that this follows simply from that very fact that they are rational beings by nature. The rational and reflective capacities that demand that we regard ourselves as free come with a normally functioning human being's nature and they require that we regard ourselves as free because of the reflective distance we have to cover in order to act: because we are reflective rational beings, we act for reasons and thus must think before we can carry out actions.²⁹¹ Korsgaardian constitutivism therefore *seems* to disagree with Ferrero that agency is not ontologically necessary. But this does not mean that the Korsgaardian constitutivist is committed to the biting-the-bullet strategy with regards to the first horn of the trilemma. According to Korsgaard, agency is a process or a state one can slip in and out of, meaning that it takes work to become or continue to be an agent, but also that reflection endorsement is a process one can fail at.²⁹² So while anthropological necessity explains the necessity of the prerequisites of agency, it does not follow that agency itself is so strictly necessary

²⁸⁸ There are exceptions, of course; human beings may have difficulty using their rational capacities to such an extent that they can establish whether something passes the reflective endorsement test, or they may lack rational capacities altogether. I will briefly go into the implications of this when discussing the third horn of the trilemma.

²⁸⁹ 'Non-rational' here means that it does not, at first sight, require rational capacities to chase one's desires – one might be able to do so unthinkingly. This is not to say that trying to satisfy one's desires is *irrational*.

²⁹⁰ Which, according to my reconstruction of Korsgaard's analysis of reflection, means that even desire-satisfaction requires that we regard ourselves as free, which introduces the need for substantive claims about autonomy, laws, et cetera.

²⁹¹ Korsgaard often makes such claims (SC 141; CA 105, for example), but she also makes it clear that we need not always be (consciously) guided by constitutive principles (SN 236f). And from theory of identity, which I will briefly discuss in chapter 3, it may follow that once one has one's practical identities in order, one might be able to engage in actions without having to reflect on them beforehand on every single occasion. This claim, however, would require more argumentative work than I can give it here. I can therefore only point it out as a possibility.

²⁹² See, for example, SC 214.

that human beings would always be agents. The reason is simple: agency, on the Korsgaardian conception discussed in the first chapter, should be conceived constructively as something that arises only through successful actions. The extent to which we are agents is therefore identical to the extent that our actions are successful and thus in conformity with the principle of our will, the CI.²⁹³ Hence Korsgaardian constitutivism can and must avoid the first horn, because it cannot and does not accept the claim that all human beings are necessarily agents. It avoids this horn by maintaining that the prerequisites for agency come with our nature, but the extent to which each individual human being constitutes herself as an agent depends on her actions and thus agency is not ontologically necessary.

At this point, one might object that if normally functioning human beings engage in reflective processes and if the process of reflective endorsement cannot but lead to agency, then all those human beings would (always) be agents. In other words, Korsgaardian constitutivists have to do more to explain how reflective endorsement may fail. Without such an explanation, my interpretation of Korsgaardian constitutivism would be as vulnerable to the first horn of the trilemma as Ferrero's theory. For this reason, the 'if properly carried out' or 'if successful' clause is necessary, because it allows for the fact that the reflective process can fail. This is important because there are many ways in which the reflective process may turn out unsuccessful: it may be interrupted, one may wrongfully judge that the law one (thinks one) makes will apply similarly in future circumstances, and there are many more such scenarios.²⁹⁴ An important addition is that the human condition requires people to at least make an effort to succeed, because reflective success (that is, the production of reasons) is what they need in order to act. And here we also find two of the reasons why one cannot reflectively decide to be a shmagent: if the process is successful, then one constitutes oneself as an agent, not a shmagent, and the maxim of using the reflective process in order to become a shmagent is unlikely to pass the reflective endorsement test. For one thing, if one were to universalize such a maxim, it would mean that everyone would become a shmagent and would thus be unable to act, even though that is what human beings are 'condemned to'. Consequently, (my interpretation of) Korsgaardian constitutivism is not committed to the view that every human being is an agent, nor is it vulnerable to the suggestion that unsuccessful deliberation is a case in point for Enoch's shmagency objection.

The second horn can be avoided too, because Korsgaardian constitutivism (on my interpretation and contrary to Ferrero's theory) does not claim that agency is something one must care about. As I have tried to argue in the previous sections, the need to attempt to constitute oneself as an agency is inescapable because even such things as the satisfaction of one's desires require agential capacities. To say that such a need is not only inescapable, but also something one should *care* about is simply one thought too many. Korsgaard, for one, advocates the view that agency is constituted by

²⁹³ Yet one continues to be accountable even if one fails at constituting oneself as an agent, Korsgaard argues, because it is a human being's 'plight' to constitute her agency and practical identities through her actions (see, for example, SC 175). This analysis of the human condition, I believe, lies at the very heart of Korsgaard's theory, because it allows her to explain why one may hold people accountable even if they fail at being an agent or even if they have come to believe that they are not free or if they have become apathetic to their freedom and their need to act. Regardless of what happens or what people believe, they will have to continue trying to build an identity (and thus constitute themselves as an agent), because that is their human condition.

²⁹⁴ Cf. SC 214.

our actions, but this does not commit her to the idea that we therefore somehow also care or ought to care about agency as such. The Korsgaardian point I have tried to make against the shmagency objection is that shmagency is not a viable alternative to agency and that constitutivism need not claim that people ought to care about agency. Thus it avoids the second horn because that misses the point. Constitutivism can argue against supposed shmagents that they are in fact showing agential capabilities and there is no need for constitutivism to argue in a further step that they should also care about their agency.

The third horn of the trilemma takes a slightly different form in relation to Korsgaard's theory. The problem here is not that there might be a large group of people who do not care about agency. Again, this would not be a problem, because Korsgaardian constitutivism does not entail that it is necessary that people care about agency in itself. But there might be people who do not have rational capacities and thus cannot possibly be agents. This group would consist of people with intellectual disabilities, for instance, and thus it would be quite small. Where the trilemma's third horn is concerned, then, the Korsgaardian constitutivism about agency I have put forth can simply bite the bullet and accept that there will be a small group of human beings who are not, and cannot be, agents. Because of their special circumstances, constitutivism may have to claim that these people cannot be fully autonomous and efficacious, but this is not a knock-down problem for constitutivism. Korsgaardian constitutivism, on my anthropological-necessity interpretation, starts from very basic reflective capacities and there are not many people who lack those and to whom one would still ordinarily describe agency.²⁹⁵

It is along these lines that I believe Korsgaardian constitutivism can circumvent all three horns of the shmagency trilemma. Admittedly, the theory of anthropological necessity I have proposed is not quite as direct with regards to the necessity of agency as Ferrero's theory, but it is better equipped to deal with the shmagency trilemma. And it is a better way of refuting Enoch's shmagency objection. On my anthropological necessity-interpretation, there is simply no need for Korsgaard to commit herself to the claim that one should necessarily care about agency.

Skewed and unfair debates

Rather than restating my case, however, I would like to take this opportunity to return to a point at which I have already hinted at the start of this chapter: something seems to be going awry in the shmagency debate.²⁹⁶ In fact, there are two distinct yet related things that have led to what I believe to be a somewhat skewed debate. First, as discussed at the end of section 2.3, there is the assumption that shmagency is a viable alternative to agency. Although this is of course a legitimate argumentative position that cannot be ruled out beforehand, it must be supported by positive evidence, which Enoch does not deliver. And this leads to a second, related issue: on the more abstract level, the burden of

²⁹⁵ For one thing, if agency is 'autonomous efficacy', then one would not expect ascribe agency to people whose reflective capacities have been impaired by Alzheimer's, for example, which is to say that there is only a very small group of people who should be called agents yet who, according to Korsgaardian constitutivism, are not agents.

²⁹⁶ I thank Joel Anderson for pressing me on this point.

proof is shifted to constitutivists in what I think is an insufficiently justified way. Rather than providing a full-fledged discussion of (the possibility of) shmagency, Enoch seems to simply assume that there could be something sufficiently like agency which lacks the constitutive goals of agency. Again, this is a legitimate line of argument, but only if it is supported by evidence. In the shmagency debate, the assumption seems to be that shmagency is possible and that that somehow means that the burden of proof lies with constitutivism. What I have been trying to do in this chapter is to give a charitable reading of the shmagency objection, investigating some of the ways in which it might pose a real problem for constitutivism. From Enoch's articles alone, however, shmagency is not sufficiently supported as an actual, viable alternative to agency and thus one could wonder where the burden of proof really lies.

A further issue with the shmagency debate is that it relies on an unfair depiction of constitutivism. As I have argued in section 2.1, Enoch's shmagency objection asserts that a to-be-shmagent can ask what it is to him how someone else characterizes his behavior and that shmagents do not care about agency. Strictly speaking, these two claims do not counter any constitutivist point. In both chapter 1 and the current chapter I have tried to show that Korsgaard's constitutivism does not involve the claim that one should care about agency as such, nor that second- or third-person claims about agency matter. Enoch's shmagent is absolutely right in asking what it is to him what someone else calls his behavior, but this misses the constitutivist point. Perhaps more so than most other moral theories, constitutivism centers around autonomy and choosing one's own principles, and thus constitutivists will agree that third-person judgments about one's behavior are completely irrelevant to the status of one's actions and agency.

Additionally, there is a very simple, straightforward constitutivist answer to the shmagency objection that I (nor Enoch) have not yet discussed. If agency relates to shmagency as a house does to a 'shmouse' (something a lot like a house, but without being built with the constitutive principles of a house in mind), then according to constitutivism, shmagency is simply agency but poorly carried out.²⁹⁷ This means that if shmagency is supposed to be a successful counterargument to, for example, Korsgaard's constitutivism, shmagency-proponents would have to argue for one of the possibilities. First, they could try to show why that analysis is incorrect and why shmagency is more than failed agency. Alternatively, they could try to argue that shmagency is indeed failed agency and this raises questions about the necessity of agency. In either case shmagency-proponents would have to clarify why shmagency is a truly distinct, stable and viable point of view with which the necessity of agency could be called into question. Furthermore, they would have to show why agency and shmagency are on level terms and why or how there could be such a thing as a moment at which a person is called to choose between those supposed alternatives. This, however, would mean that the debate has to take a drastic turn and move from discussing the necessity of agency to the possibility of shmagency. Only after the possibility of shmagency has been shown can shmagency be used to counter the constitutivist line of thinking I have put forth in this paragraph.

²⁹⁷ Cf. SC 29. On the Korsgaardian understanding of action and agency, this might be the result of acting on maxims that do not sufficiently render the actor autonomous. This, of course, raises the question if autonomy comes in degrees and, if so, where the threshold lies, but these are questions I cannot answer here.

All this is not to say that constitutivism is right on all points. In fact, I have used this chapter to show that shmagency could be a real problem for constitutivism and that constitutivism therefore has to provide an answer to the shmagency objection, with my anthropological-necessity interpretation of Korsgaard's work as only one of several possible lines of thought. But the debate would surely benefit from a better understanding of constitutivism in all its varieties, and a corresponding assessment of its strengths and weaknesses. Furthermore, these considerations do not prove that the burden of proof lies with the advocates of shmagency and that my constitutivist proposal is without problems. However, they do show that the shmagency objection might have been more powerful – and the debate more fruitful – if its advocates had done more work to provide a level playing field.

Chapter 3: Necessity and normativity

In the previous chapters I have outlined some of the key elements of Korsgaard's writings on constitutivism, agency, action and reflection. I have subsequently exposed her theory to what I believe to be a potentially devastating critique: David Enoch's shmagency objection. 'Potentially' is a crucial adverb here, because as I have argued in the previous chapter, Enoch casts the critique in such a form that it does not touch Korsgaard's constitutivism in any direct way. It might be interpreted differently, as I have also argued, but I have not found an interpretation that would bring my version of Korsgaardian constitutivism in difficulty. The anthropological-necessity interpretation I have outlined in order to circumvent Enoch's objection does, however, have some distinct problems of its own. This final chapter will comprise a discussion of some of these problems and normative costs.

The chapter will have the following structure. First, I will discuss William FitzPatrick's critique of Korsgaard's constitutivism. FitzPatrick provides an argument to show that Korsgaardian constitutivism in fact does too much: where a more modest theory of the generalizability of reasons might have sufficed, FitzPatrick argues, Korsgaard takes a few steps too many by trying to show that reasons are subject to the CI or by trying to extend hypothetical imperatives about what one ought to do to more general, categorical imperatives. Just how this critique works and why in the end it fails will be the subject of the first section. Section two will consist of a short discussion of Sergio Tenenbaum's estrangement argument, which – roughly put – goes something like this: Korsgaard's constitutivism, through its focus on categorical imperatives, leaves agents estranged from the particular ends they intend to follow. The final section will comprise a discussion of the relation between (moral) autonomy and personal identity. I will there again take my cue from FitzPatrick, who claims that constitutivism has anything but shown that morality and personal identity are subject to the same principles.²⁹⁸ This is of course a central element of Korsgaard's recent work, so FitzPatrick's claim deserves exploration and discussion, for it might point to an important shortcoming in Korsgaard's work.

3.1 Overstating the importance of universal validity

FitzPatrick criticizes Korsgaard's work in a number of ways.²⁹⁹ As a first approximation, FitzPatrick's two points of critique are as follows. The first is aimed at Korsgaard's supposed attempt to argue, first, that when one wills something, one is subject to hypothetical imperatives and, second, that these hypothetical imperatives should also be extended into categorical imperatives. This extension, according to FitzPatrick, is deeply problematic. The second type of criticism revolves around the idea that Korsgaard may be overstating the importance of the universal applicability of reasons: it might have sufficed for her to argue that one may assume that a reason one has for doing X in certain circumstances will also be a reason for someone else to do X in similar circumstances. To claim that reasons should be potential universal laws, then, would be taking the point too far.

²⁹⁸ William J. FitzPatrick, "How not," 44.

²⁹⁹ I will discuss two types of critique here and one in 3.3.

Extending hypothetical imperatives

Both counterarguments target important elements of the Korsgaardian theory I have described in chapter 1 and defended in chapter 2. To start with the first objection, FitzPatrick's idea is that hypothetical imperatives could be a somewhat unproblematic way of showing the usefulness of imperative-based theories.³⁰⁰ More precisely put, he appears to find that Korsgaard fails to show how the normativity of these supposedly easier to understand and less controversial hypothetical imperatives works, even though showing that would be easier than showing how and why one should be subject to the categorical imperative.³⁰¹ The more important point, however, is that one could grant Korsgaard whatever theory she has of hypothetical imperatives, and that this still would not show how *categorical* imperatives work. The question thus is whether any model of hypothetical normativity could be extended to moral normativity.³⁰² Perhaps most concretely put, FitzPatrick asks 'whether Korsgaard has provided any new and compelling arguments' to explain how the CI works.³⁰³ The core of the objection, I therefore believe, is that (a) Korsgaard uses the CI as a more general and perhaps more directly moral extension of the less controversial hypothetical imperative, even though (b) she fails to explain the normativity of the hypothetical imperatives and (c) she also does not provide adequate arguments to show that the CI can be derived from the hypothetical imperative.

In answering to Enoch's challenge, I have tried to argue that rational beings must regard themselves as free and autonomous and thus must choose the CI as the principle of their will. If FitzPatrick's objection successfully shows that it is unclear how the CI's normativity is supposed to work, then this puts Korsgaardian constitutivism in an awkward position. That is, if Korsgaard has not explained how the CI can be derived from hypothetical imperatives and how the CI's normativity works, then the first part of the claim that the CI is constitutive of agency (one of Korsgaard's key points) is deeply problematic.

However, there is a quite simple problem with FitzPatrick's critique: to claim that the CI is supposed to be an extension of the hypothetical imperative is to misunderstand Korsgaard's theory. If anything, the opposite is true: as I have tried to show in chapter 1, Korsgaard thinks that hypothetical imperatives are constitutive of agency because autonomy is, and autonomy is useless without efficacy, roughly put. Additionally, as I have also discussed before, Korsgaard claims that the hypothetical and categorical imperative may be inseparably combined, but she also states that they fulfill different roles.³⁰⁴ The claim that the CI must be derived or argued for by 'extending' the hypothetical imperative is quite simply false, which means that FitzPatrick wrongly shifts the burden of proof to Korsgaard, who is subsequently supposed to give an account of the normativity of the CI because it does not follow from hypothetical imperatives.³⁰⁵ One can of course justifiably ask how the CI's normativity works or

³⁰⁰ See FitzPatrick, "How not," 45-52.

³⁰¹ See, for example, *ibid.*, 53.

³⁰² *Ibid.*, 52.

³⁰³ *Ibid.* Cf. William J. FitzPatrick, "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity," *Ethics* 115, no. 4 (July 1, 2005): 651-691, 690.

³⁰⁴ To give only one example, she writes that 'the hypothetical imperative that tells you that if you will the end, you must will the means to that end, and the categorical imperative that tells you that you must will your maxim as a universal law.' (SC 58).

³⁰⁵ See, for example, FitzPatrick, "How not," 52.

how it is to be argued for, but only after an accurate representation of Korsgaard's views. Because it is not true that Korsgaard's theory relies on the possibility of extending the model of hypothetical imperatives to the CI, the burden of proof with regards to these objections relating to the CI lies with FitzPatrick.

The universality of reasons

The second type of critique FitzPatrick gives is that Korsgaard reads too much into the universal status of reasons. FitzPatrick points out that it may be true that our will needs principles, but that this does not imply that our will needs the CI; it may suffice that our reasons for action apply similarly in like situations.³⁰⁶ Thus in order to avoid particularistic willing or something like a solipsism about reasons, the weak universality requirement just described might do the trick. All Korsgaard needs to accept to avoid particularism is the following, which she also discusses herself: 'if I have a reason to do action-A in circumstances-C, then I must be able to grant that you also would have a reason to do action-A were you in circumstances-C'.³⁰⁷ This is a more modest position to take than Korsgaard's claim that reasons must be potential universal laws. FitzPatrick therefore concludes that it is a stretch to claim that reasons must be (potential) universal laws and that agents must therefore follow Kant's CI.³⁰⁸

If correct, then FitzPatrick's objection poses a serious problem for the Korsgaardian theory of agency I have been defending. If it is true that the will needs no more than a weak universality requirement and thus does not need the principle of autonomy (the CI), then much of what I have said in response to Enoch is false. To give just one example, it would be wrong to say that we must, because of our reflective nature, ultimately accept the CI as the principle of our will, for the weaker universality requirement would have sufficed to account for the fact that human beings need reasons for action. And this weaker requirement, I believe, does not rule out shmagency in any clear or direct way, which would mean that constitutivism is back where it started.

FitzPatrick is right to say that the argument he picks out leaves a lot to be desired.³⁰⁹ He focuses on Korsgaard's argument against particularistic willing as given in *Self-Constitution*, chapter 4, and in isolation this argument is indeed not very convincing. Korsgaard there simply argues that particularistic willing is impossible and she does very little to really show that there is only one alternative: following the CI.³¹⁰ But Korsgaard elsewhere provides what is missing here. Again, in *Self-Constitution* she argues that particularistic willing is impossible. This is the first step. Next, the will therefore needs some sort of overarching principle, whether it be FitzPatrick's weaker universality requirement, the CI or something else. Now if the will is to be free, there is a further requirement: whatever law or principle the will has must be its own.³¹¹ Nothing outside the will can determine what that principle is, and thus the only strict requirements are that it is a law or principle and that this is the will's own, rather than

³⁰⁶ Ibid., 57.

³⁰⁷ SC 191.

³⁰⁸ FitzPatrick, "How not," 57.

³⁰⁹ See, for example, ibid., 59.

³¹⁰ See SC 72-76.

³¹¹ Cf. SN 98.

something that is imposed on it. And that, as I have discussed before, is precisely what Kant's CI also states. That, I believe, is why Korsgaard claims that the CI is the principle of a free will.³¹²

If Korsgaard's argument is correct, the CI is already present as the principle of the autonomous will. FitzPatrick then misses the point when he wonders why an agent should follow Kant's CI.³¹³ Korsgaard does not advocate the view that people somehow already have a will and that they should be convinced to accept Kant's CI – out of a manifold of moral theories and principles – as their principle of choice. Insofar as people have a free will or work on constructing one, they already (probably unknowingly) accept the CI as their principle, because the CI *is* the principle of a free will.³¹⁴ The CI, according to Korsgaard's constitutivism, is not just one among many possible principles available to the will and therefore it would be a mistake to ask why the CI should be chosen over those other principles. To conclude, FitzPatrick is right in saying that Korsgaard's argument from *Self-Constitution*, chapter 4, leaves a lot to be desired. But she fills these argumentative gaps elsewhere and thus FitzPatrick's counterargument leaves out important elements of her theory. It is therefore incomplete at best. This means that, at least where FitzPatrick's charges about the universality of reasons are concerned, my Korsgaardian response to Enoch is in the clear. Although Korsgaard's argument against particularistic willing by itself does not suffice to show why agents must choose the CI as their principle (rather than just accept a weak form of universality), there are other ways of arriving at this conclusion, which Korsgaard investigates in other writings. In their current form, then, FitzPatrick's counterarguments do not harm the theory of agency I have proposed in the previous chapters. They do, however, force those taking an interest in Korsgaard's constitutivism to look closer at the different elements of the theory, such as the status of the CI, and to better understand or explain their connections. FitzPatrick's objections show that constitutivists have either unconvincingly or unclearly argued for their theses, meriting a critical review of both FitzPatrick's and constitutivism's points. Still, even if (or precisely if) FitzPatrick's critique is adequately dealt with and it is true that the CI is the principle of a free will, there might be further normative costs, which I will discuss in the next section.

3.2 The Categorical Imperative and estrangement

Korsgaard's constitutivism is meant to show that morality and personal identity are deeply interwoven.³¹⁵ Sergio Tenenbaum has recently expressed doubts about this ambition, arguing that, if anything, constitutivism achieves the opposite: it alienates people from their individual goals and projects.³¹⁶ Constitutivism, Tenenbaum writes, 'forge[s] a commitment behind the back of the agent. [It] puts us in a position in which what we immediately take to have value gives way to a norm that is

³¹² See, for example, SN 98.

³¹³ FitzPatrick, "How not," 58.

³¹⁴ Alternatively, people gain a will by meeting the constitutive standards of agency, Korsgaard states (SC 70).

³¹⁵ The alienation objection I will discuss here is part of a large body of research on morality and estrangement. I discuss only Tenenbaum's critique, because that is tailored specifically to (Korsgaard's) constitutivism.

³¹⁶ Sergio Tenenbaum, "Formalism and Constitutivism in Kantian Practical Philosophy," *Philosophical Explorations* 22, no. 2 (May 4, 2019): 163-176.

not grounded on the direct object of our will'.³¹⁷ The point is that agency, according to most constitutivists, is defeasible, meaning there are different degrees of success at being an agent which may impact the extent to which one counts as an agent. Assuming that one's agency is an overarching end, then, Tenenbaum argues that people have numerous subordinate, individual ends that might be at odds with the end of being an agent. And because one's agential status is defeasible, it should in principle be possible to let those subordinate ends take precedence over agency, which might not be an end people actively pursue or embrace.³¹⁸ Tenenbaum summarizes his critique as follows:

If I take an interest in being an artist, the constitutive conditions of being an artist flow from the object of my will, while the constitutive conditions of agency are those that I find myself committed [to, B.G.] by the inescapability of agency. While one is foregrounded as the object that engages my will, the other seems to be at best a secondary commitment.³¹⁹

Tenenbaum thus seems willing to grant that agency and efficacy are constitutive of agency and that the CI is the principle of autonomy, and argues that, even then, constitutivism is not exactly an appealing theory. Although Tenenbaum presents this simply as an alienation objection, he in fact seems to be putting forth a two-layered critique. First, he could be read as arguing that constitutivists have failed to clarify how the hierarchy of commitments works. Second, he might be read as pointing out that even if constitutivism successfully explains said hierarchy, it will still be unattractive, because it will likely give agential commitments priority over other commitments and this, supposedly, leaves people alienated from what they really care about. That is, if constitutivism maintains that one should value agential commitments over whatever other commitments one has, then this will alienate people from those latter commitments that in fact are quite important to them.

With regards to the first point of critique, I believe one possible Korsgaardian answer is this. There is no real conflict or hierarchy between agential and other, 'subordinate' commitments or ends.³²⁰ According to the picture I have sketched in chapter 1, we must set ourselves ends autonomously if they are to count as truly *our* ends. If we want to truly set ourselves ends, their corresponding maxims must pass the reflective endorsement test; we must be able to will them as laws. As said in the previous section, that is the only requirement for the ends and principles we choose. Constitutivism does not rule out any specific ends beforehand because they might be at odds with the demands of agency. In fact, constitutivism holds that by deciding what we ourselves want to do and value we also constitute ourselves as agents.³²¹ Agency is not some further and supremely

³¹⁷ Ibid., 168.

³¹⁸ Ibid., 167-168.

³¹⁹ Ibid., 168. Cf. Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy & Public Affairs* 13, no. 2 (1984): 134-171, 136-138.

³²⁰ That is to say, there is no predetermined hierarchy, but there can be conflicts one must settle by attempting to unify oneself as much as possible, so that one can be the cause of one's ends.

³²¹ Making value dependent upon a conscious act by a person may lead to a new estrangement worry, according to which people might feel estranged because the world they live in is essentially a blank slate. This, however, is simply the price of avoiding realism about value (or at least of accepting constructivism or constitutivism about value).

valuable end or constraint that we must keep in the back of our minds whenever we decide what we have reason to do, nor is it something the achievement of which requires a separate effort.³²² When one engages in reflective endorsement, one establishes what one has reason to do and thus which practical identity must outweigh potentially conflicting identities. By doing so, one works on one's own project but also simultaneously and unavoidably constitutes oneself as an agent. Agency is simply the result of a person autonomously settling what she values and therefore there is no real conflict. Constitutivism might be interpreted either as saying that agency is a precondition for setting oneself any values at all, or that it is the 'byproduct' of the process of doing so, but either way, agency is not in a direct conflict with other ends one might set oneself.³²³

Tenenbaum's second point can be refuted quite easily. He is right in thinking that agency has a special status according to constitutivism. But on Korsgaard's view, this does not mean that agency alienates people from what they truly care about. To take Tenenbaum's example of being an artist, such a practical identity provides you with reasons for action. Practical identities play a crucial role in the process of establishing what one has reason to do and thus they are essential to autonomy and, ultimately, agency.³²⁴ If you reflectively decide that you want to be a musician and that therefore you have reason to practice playing the piano, for instance, then you also constitute yourself as an agent. Having practical identities, valuing them and working out what reasons they give you for action is crucial for constituting oneself as an agent, so if anything, constitutivism would urge people to create practical identities for themselves.³²⁵ There is, however, the caveat that the reasons one's practical identity give must pass the CI tests, so it is not as if anything goes. Being a musician will give you reason to practice the piano as much as possible, but doing so even when people around you are in immediate danger would probably be ruled out by the CI. This does not mean that constitutivism alienates our to-be-pianist from what she cares about the most. It simply means that there may be circumstances in which she will be called upon to see that her practical identity as a human being provides her with reasons that outweigh her artistic reasons for the time being. It is therefore quite simply false to state that constitutivism places supreme value on agency and that this impacts all other possible practical identities one might choose. Constitutivism encourages people to actively engage in choosing and building practical identities of their own, because these are crucial to building one's agency and to safeguarding one's autonomy.

³²² The only constraints Korsgaardian constitutivism seems to impose are these: one must be able to will one's principles as laws and these must be one's own (freely chosen) principles. Since agency is constituted by one's actions and thus does not exist prior to choice and action, it cannot feature as something normative to be kept in the back of one's mind in the reflective process.

³²³ Korsgaard does write that, before we can value our particular practical identities, we must value our 'human identity', which should not be mistaken for agency and which means no more than that we must endorse our human need to have an identity of our own (SN 125). This is a basic statement about the preconditions of valuing one's practical identities, which does not commit Korsgaard to the view that morality must always or generally outweigh other, perhaps more personal practical identities.

³²⁴ Cf. SC 214.

³²⁵ Cf. SN 129-30.

3.3 Autonomy and identity: moral saints

In the preceding sections I have tried to show why practical identities and individual commitments and values are essential to Korsgaard's constitutivism, contrary to what Tenenbaum believes. Following Korsgaard – if my interpretation is correct – I have also repeatedly stressed that our reasons for action must conform to the CI. So while practical identities are hugely important, there is an important proviso: whatever reasons our practical identities provide us with must withstand reflective scrutiny and thus pass the CI test. In this last section I will describe two important worries about this constitutivist approach to personal identity. The first is inspired by Susan Wolf's "Moral Saints", in that there might be good reason to wonder if Korsgaardian constitutivism does not require that we behave *too* morally.³²⁶ Although there is some overlap between this type of objection and the alienation objections discussed in the previous section, Wolf's 'moral saints' objection, I believe, is better understood as a demandingness objection than an estrangement objection.³²⁷ The second worry is voiced by FitzPatrick, who points out that constitutivism at best ends up with a very odd notion of moral goodness.³²⁸

Before discussing the first potential problem, it might help to briefly consider Korsgaard's theory of moral goodness and personal identity a bit more directly. According to Korsgaard, actions done properly (i.e., good actions) are right actions (i.e., morally good actions).³²⁹ In fact, Korsgaardian constitutivism is the view that something can only truly be X if it lives up to X's constitutive standards, meaning that an action is truly an action only if the agent carrying it out adheres to the constitutive principles of action and thus to the CI. As a result, a real action, properly carried out, is such that it could be willed as a universal law. A real action has the right form – the right relation between the act and the reason for which it is done. The form of an action is what enables an action to be willed as a universal law. If the form is right, the action in question meets the constitutive principles of action and therefore it is a good action. For that same reason, if my interpretation of Korsgaard is correct, when an action meets the constitutive standards of action, it is more than mere behavior. And by meeting those standards, it is not only truly an action but also a morally right action.

Korsgaard's theory of identity works along similar lines. '[I]f you are good at being a person, then you'll be a good person', Korsgaard writes.³³⁰ Why? For Korsgaard, being good at being a person means having the right 'constitution'. This has been implicitly present in the previous chapter, where I discussed the different interpretations of shmagency. Someone who merely chases after whatever desire he happens to have is not unified, Korsgaard argues.³³¹ Such a person fails to make up his mind and establish which actions are expressive of himself, and instead simply does what his 'gut' tells him to. When someone engages in deliberation, on the other hand, she 'imposes unity' on herself (or at

³²⁶ This is the worry Susan Wolf voices (albeit not with regards to Korsgaard's theory specifically) ("Moral Saints," *The Journal of Philosophy* 79, no. 8 (1982): 419-439).

³²⁷ Perhaps it was originally meant as an alienation objection, but in the present context it has interesting merits when interpreted as a demandingness objection.

³²⁸ FitzPatrick, "How not," 44 and 55-56.

³²⁹ See, for example, SC 160-161.

³³⁰ SC 214. Cf. SC 176.

³³¹ SC 177.

least attempts to do so), as Korsgaard calls it.³³² Through the deliberative process I have described, a person decides what matters to her – that is a crucial element of the deliberative process – and by doing so organizes her thinking, feelings et cetera into a unified whole which can then be the cause of an action, for which she is responsible and which can be attributed to her as a person.³³³ The idea thus is that whatever else we are doing when we engage in deliberation, we are also settling what matters most to us as a person and thus we are organizing or creating unity within ourselves. And when people succeed in being unified, they can be the autonomous and efficacious cause of actions, which can be willed as universal laws and are thus morally good. Hence Korsgaard’s conclusion that being good at being a person means being a good person.

Now to return to the potential normative costs of this theory, the first worry, again, is that in Korsgaard’s theory the CI might be supposed to guide an agent’s entire life and all her actions and that that is a bad thing. If all the reasons for everything agents do must conform to the CI, does that not mean that these agents (must) become rather like moral monks, with their lives fully devoted to the demands of morality? After all, that is how Wolf defines moral sainthood: to let your motivation to be as morally good as possible be the core determinant of your character.³³⁴ There are at least two ways of answering to this normative objection. The first and less interesting one is this: according to constitutivism, morality (the CI, more specifically) is only binding because it is the principle of your own free will – a principle everyone must adopt autonomously. Morality thus does not exist prior to or outside of one’s will but is instead constructively understood as arising only through the employment of reason. Hence there is no moral domain or demand outside of oneself to which one could devote oneself. The second possible answer is more interesting. Wolf argues that a moral saint could probably not defend taking an interest in gourmet cooking, because the resources spent at that could have been put to better use elsewhere.³³⁵ As I have argued before, however, this picture does not apply to constitutivism. It is essential that people have personal interests, preferences, identities and so on. By actively shaping those, we become people with identities of our own and can therefore constitute ourselves as persons. And by trying to make ourselves into persons, we become better at being a person and so become better persons. Thus it is simply false that constitutivism demands that people become moral saints. Their moral improvement is not a goal in itself to which they must devote themselves, but a necessary byproduct of people’s striving to be a person. The goal of this self-constitution is to shape and become who one *wants* rather than *happens* to be. Although moral improvement is inseparably linked to self-improvement,³³⁶ constitutivism does not claim that people should be moral saints.

This answer may put to rest any moral-saints worries. But it also opens the door to criticism voiced by FitzPatrick, who points out that Korsgaard’s constitutivism relies on a rather unintuitive

³³² SC 179.

³³³ Ibid. The idea here, as described before, is that a person is only truly the cause of an action if she has autonomously chosen to carry it out, and that requires unity of the will.

³³⁴ Wolf, “Moral Saints,” 421.

³³⁵ Ibid., 422.

³³⁶ And self-improvement is inseparably linked to the inescapable predicament of building a ‘self’ in the first place.

notion of moral goodness.³³⁷ The worry here is that moral goodness is cashed out or defined in terms of agential unity and that that is either a change of subject – rather than an actual explanation – or that it is an incredible and underexplained coincidence that moral goodness and agential unity overlap.³³⁸ For this worry to turn into a real, non-question-begging point of critique, it helps to consider an example FitzPatrick gives: ‘it would seem that one could “successfully act” in effectively detonating a bomb in a marketplace no less than in effectively helping a neighbor cross the street, yet the former is not *good action* in any sense relevant to ethics’.³³⁹ In other words, it might be possible that the act of detonating a bomb in a marketplace meets the constitutive standards of action and that Korsgaard’s constitutivism would therefore deem it a successful and thus morally good action, even though it evidently is not. So if a unified agent could kill a crowd of innocent civilians,³⁴⁰ then clearly agential unity is not identical to moral goodness.

Equally clearly, there is something missing in FitzPatrick’s example: the reason behind the act. ‘Detonating a bomb’ is not a description of an action, but of an act. As seen before, human beings need reasons for action and these need to pass the reflective endorsement test. And it is highly doubtful whether there is any motive that could sufficiently justify killing innocent shoppers for the maxim to pass said test. I believe that to be self-explanatory enough, but will substantiate it nonetheless. Suppose the reason for detonating the bomb is that one wants to go to heaven. This would not pass the practical contradiction test, because if one were to universalize the action of detonating a bomb to go to heaven, no one would be going outside anymore, for fear of being blown up by someone else. So as not to dwell on this any further, suffice it to say that FitzPatrick’s example is not what it is intended to be. I share his view that cashing out moral goodness in terms of agential unity is an unintuitive move, but unintuitive is something quite distinct from problematic or wrong. Korsgaard has good arguments to back up said connection and in the absence of any compelling counterarguments it seems better to just follow the argument to where it leads.³⁴¹

There is one further potential issue with identifying moral goodness with agential unity. One may wonder whether such a position does not commit one to conceding that people who suffer from weakness of the will (or people who otherwise lack the abilities required to unify themselves as agents) are incapable of agential unity and thus of moral goodness. Are such people not perfectly capable of doing morally right things? And is it not harsh to claim that they are not unified and that they therefore simply cannot be morally good? To answer this objection, we should distinguish between moral actions and behavior that has morally desirable results. Korsgaard’s constitutivist starting point is that in order to know what a (morally) good action is, we need to know what an action is in the first place. As I have discussed, this leads her to ultimately conclude that a good action has a maxim that one can will as a

³³⁷ FitzPatrick, “How not,” 44 and 55-56.

³³⁸ FitzPatrick does not clarify which of these two he considers to be the real problem (see *ibid.*, 44).

³³⁹ *Ibid.*, 56. Italics in original.

³⁴⁰ Or if someone could become a unified agent by doing so (FitzPatrick is not entirely clear on this).

³⁴¹ Additionally, it remains an open question if Korsgaard really thinks (or should think) that moral goodness is (and only is) agential unity. Perhaps moral goodness and agential unity are two sides of the same coin, because either (i) moral goodness and agential unity are different ways of describing what makes an action good or (ii) moral goodness and agential unity are two distinct but inseparably combined and necessary outcomes of the reflective process.

universal law and that constitutes the actor as an agent. Such actions, as said, require a reflective process through which one unifies oneself, and insofar as one is incapable of unifying oneself, one is indeed incapable of actions in this Korsgaardian sense. Still, even those who cannot unify themselves can of course do a morally desirable thing X. But Korsgaardian constitutivism, if my interpretation is correct, holds that insofar as one is not the autonomous, efficacious entity that decided to set oneself the end of X, one is also not the cause of X. And if one is not the cause of X, one cannot be praised for doing X. Hence, Korsgaardian constitutivism, I believe, will have to bite the bullet here and concede that there may be people who are not capable of agential unity. Such people then cannot cause actions and therefore, again, they cannot be praised for what they do, meaning that they can indeed not be said to be morally good. Being morally good is a label one deserves by doing good things, and if one, strictly speaking, does not cause any (good) actions, one *does* nothing and thus Korsgaardian constitutivism must here judge that such a person is indeed not morally good.³⁴²

To conclude, the Korsgaardian theory I have outlined and defended seems to have several (normative) costs or problems, but it is well-equipped to deal with those. Whether it be FitzPatrick's charges against the supposed move from hypothetical to categorical imperatives, Tenenbaum's critique regarding estrangement from personal projects, or the two types of problems I have discussed in this section, I have tried to argue that there are good reasons to believe that constitutivism can deal with those issues. In fact, many turned out not to apply to Korsgaardian constitutivism in the first place and thus I will end this chapter in a similar fashion to the previous one: just as on the meta-ethical level, debates around Korsgaard's theory of self-constitution on the level of normative ethics would benefit from a deeper understanding both of the different elements of Korsgaard's theory and of the way they are connected. Without that, there may be plenty of objections and worries to be found, but these have only superficial philosophical value. As I have tried to show in this chapter, on closer inspection they often turn out to be targeting strawman versions of constitutivism and thus any possible real normative costs of Korsgaard's theory have yet to be established.

³⁴² I intentionally leave open the question if this means that such a person is morally bad. It seems to me that that is a very hasty conclusion.

Conclusion

Korsgaard's constitutivist theory has a number of merits, which I hope to have shown at least in passing. One, I believe, is her theory of personal identity, to which I will return shortly. Another is the fact that her theory is built on quite modest foundations: rational nature and reflection. Still, it comes with plenty of unclarities and problems. There is one problem in particular which Korsgaard and Korsgaardians have not previously ventured to deal with: Enoch's shmagency objection. In this thesis I have attempted to take up this glove, using as my guiding research question how Korsgaard's constitutivism can counter Enoch's shmagency objection. I have divided this question into three sub-questions, corresponding with the three chapters that have comprised this thesis:

1. What elements constitute Korsgaard's theory and what is their connection?
2. How can those elements be utilized to refute Enoch's objection?
3. What are the potential normative costs of this way of answering to Enoch's critique?

As for the first (sub-)question, I can only give a rough sketch here. The key to the Korsgaardian system I have tried to describe is rational and reflective human nature. Korsgaard starts from the basic ideas that human beings (i) have to choose and act and (ii) are reflective beings, who must act on the basis of reasons. This combination leads Korsgaard to conclude that human beings must decide what they have reason to do and that they must therefore regard themselves as free. They must then go through the process of reflective endorsement, which means that they must establish whether the maxim of their intended act can pass the CI test and thus whether it is in accordance with the principle of their free will. If it is, then they can freely will the maxim and if it is not, then they have a duty to refrain from carrying out the action. And it is through this process that human beings can constitute themselves as agents and thus, according to Korsgaard's definition, as autonomous and efficacious beings. Through deliberation about what one can reflectively endorse, one comes to know what ends one can freely will (making one autonomous) and what means one must also will in order to achieve said ends (making one efficacious).

The answer to the first sub-question is crucial not only because it is an attempt at providing a systematic account of the core elements of Korsgaard's theory, but also because it lays a foundation for an answer to the second sub-question and to the main research question. In order to refute Enoch's critique of constitutivism's supposed inability to ground the necessity of agency, I have gone through a number of steps. First, I have tried to show that shmagency as Enoch defines it is not possible, because human beings are reflective and must therefore regard themselves as free and must therefore also engage in the process of reflective endorsement. This means that they must see which of their maxims, if any, they can will as laws and so they establish which actions will constitute them as agents. To make my point more explicit than I have made it thus far, this entails that there are in essence two outcomes to the reflective process: either it is successful and one engages in good action and thus constitutes oneself as an agent, or it is unsuccessful and one engages in bad action and thus fails to be or become an agent. Either way, shmagency as being something like, but not

identical to, agency is not possible: either one is an agent or one is bad at being an agent, but then one is not autonomous and not efficacious and thus not enough like an agent to match Enoch's definition of shmagency. Next, I have discussed why shmagency interpreted as being half-hearted or as living a life centered around the satisfaction of desires is also impossible. As before, the point is that human beings are reflective beings and that therefore no desire has direct normative force. People must decide what desire they will try to satisfy and this means regarding themselves as free and engaging in a process of establishing what they have reason to do, and this in turn means that they either constitute themselves as agents or fail to do so and, again, neither option matches a suitable definition of shmagency.

There are normative costs attached to my Korsgaardian approach, which I have discussed in chapter 3. One might worry that Korsgaard's claim that reasons must be potential universal laws leads to an overly demanding theory or to estrangement, for example. As I have tried to show, Korsgaardian constitutivism is well-equipped to deal with such challenges and worries, not least because they quite often seem to rely on a misunderstanding of Korsgaard's work. This also explains why I have generally refrained from criticizing Korsgaard's theory, choosing instead to explain and defend it. Debates around Korsgaard's work, if they can be called debates at all, tend to be a bit disappointing: as I have tried to show, many of the criticisms Korsgaardian constitutivism receives seem to be missing the point, and Korsgaard makes little effort to either refute critique or show why opponents are misunderstanding her theory. It is for this reason that I have decided not to voice more points of critique, but to take stance as a Korsgaardian – the current debate, I believe, benefits more from the latter than the former. There may be plenty of other problems with Korsgaard's theory and these deserve due attention, but this can only happen in a proper philosophical debate where all parties involved agree on the terms and take a stance, and I humbly hope to have made a small contribution to this.

As for my own position, then, it is an attempt at providing a Korsgaardian theory even though it at times differs from what Korsgaard writes. As discussed in section 2.2, she has written at least one essay in which she invites the interpretation that we do need reasons to be an agent, and I cannot establish precisely where this leaves my approach. To claim that there need *not* be reasons to be an agent seems to be more in line with the core elements of her theory that I have been discussing, but whether Korsgaard would personally agree with my theory is of course an open question. Furthermore, my theory will have flaws and weaknesses and there is one problem in particular that I must briefly discuss. I have proposed that the necessity of agency should be understood as coming from 'anthropological necessity', meaning that human beings are rational and reflective and therefore must regard themselves as free. This might be interpreted as a downright is-ought fallacy. But it is not, because the point is that it is an analysis of human nature that humans are reflective beings. The claim that they must engage in reflection and the corresponding claims about the reflective endorsement test therefore are not supposed to be read normatively, but simply as statements about the human condition and about how reflective endorsement works. A further problem is that I have not shown to have considered (and ruled out) all possible avenues for shmagency. There may be still be alternatives to shmagency being not-quite-agency, bad agency,

desire-satisfaction or half-heartedness and the possibility of there being such an alternative merits future research.

To conclude on a more personal note and return to Kant's question I have used as a motto for this thesis, if it does not lead to estrangement, then what is it that Korsgaard's theory has to offer with regards to personal identity? Who do people become if morality and identity are deeply intertwined? What, as Kant puts it, will be the result of our morally right actions?³⁴³ Morally right actions are free and autonomous actions and these constitute us as free and autonomous beings. Thus, whatever practical identity we constitute through such actions is truly our own: it is an identity we have freely chosen to build, one for which we are responsible and one of which we can be proud. Furthermore, even though I have not discussed this in much detail, Korsgaard has made a beautiful attempt to show that because every identity is real only if it is constituted by actions, there is no 'you' over and above or outside of what you do. There is no personal existence outside of what we do. No one has described the personal effect of such philosophical theorizing about personal identity and existence quite as beautifully as Derek Parfit, whom I will therefore give the final word:

When I believed that my existence was such a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air.³⁴⁴

³⁴³ R 6.

³⁴⁴ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon, 1992), 281.

Acknowledgements

This thesis would not have been what it is now without the many helpful comments and insights I have received from Joel Anderson. Many thanks for helping me throughout the master's program and, for all the insightful feedback and discussions about different versions of this thesis and simply for being an invaluable source of inspiration. Much gratefulness also goes out to Micha Werner for providing me with many considered and insightful comments. And warm thanks go out to my fellow students, friends, family and, of course, my wife, for help, support and interesting discussions.

Bibliography

- Arendt, Hannah. *The Life of the Mind*. One-Volume ed. A Harvest Book. San Diego: Harcourt, Inc, 1981.
- Berteau, Stefano. "Constitutivism and Normativity: A Qualified Defence." *Philosophical Explorations* 16, no. 1 (March 2013): 81-95.
- Beyleveld, Deryck. "Transcendental Arguments for a Categorical Imperative as Arguments from Agential Self-Understanding." In *Transcendental Arguments in Moral Theory*, edited by Jens Peter Brune, Robert Stern, and Micha H. Werner, 141-159. Berlin, Boston: De Gruyter, 2017.
- Beyleveld, Deryck, and Marcus Düwell. *The Sole Fact of Pure Reason: Kant's Quasi-Ontological Argument for the Categorical Imperative*. De Gruyter, in print.
- Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *The Philosophical Review* 115, no. 2 (2006): 169-198.
- . "Shmagency Revisited." In *New Waves in Metaethics*, edited by Michael Brady, 208-233. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan, 2011.
- Ferrero, Luca. "Constitutivism and the Inescapability of Agency." In Shafer-Landau, Russ. *Oxford Studies in Metaethics*. 4, 303-333. Oxford: Clarendon Press, 2009.
- . "Inescapability Revisited." *Manuscript* 41, no. 4 (December 2018): 113-158.
- . "The Simple Constitutivist Move." *Philosophical Explorations* 22, no. 2 (May 4, 2019): 146-162.
- FitzPatrick, William J. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity." *Ethics* 115, no. 4 (July 1, 2005): 651-691.
- . "How not to be an ethical constructivist: a critique of Korsgaard's neo-Kantian constitutivism." In Carla Bagnoli, ed., *Constructivism in Ethics*, 41-62. New York: Cambridge University Press, 2013.
- Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, no. 23 (1969): 829-839.

———. “Freedom of the Will and the Concept of a Person.” In *What Is a Person?*, edited by Michael F. Goodman, 127-144. Totowa, NJ: Humana Press, 1988.

Gewirth, Alan. *Reason and Morality*. Paperback ed. Chicago, Ill.: Univ. of Chicago Press, 1995.

———. *The Community of Rights*. Chicago: University of Chicago Press, 1996.

Kant, Immanuel. *Critique of Practical Reason*. Revised Edition. Edited by Mary J. Gregor. Cambridge; New York: Cambridge University Press, 2015.

———. *Groundwork of the Metaphysics of Morals*. Edited by Mary J. Gregor. Cambridge; New York: Cambridge University Press, 1998.

———. *The Metaphysics of Morals*. Edited by Mary J. Gregor. Cambridge; New York: Cambridge University Press, 1996.

———. *Religion within the Boundaries of Mere Reason and Other Writings*. Revised edition, Second edition. Cambridge; New York: Cambridge University Press, 2018.

Kleingeld, Pauline. “Moral Consciousness and the ‘Fact of Reason.’” In *Kant’s Critique of Practical Reason: A Critical Guide*, edited by Andrews Reath and Jens Timmermann, 55-72. Cambridge University Press, 2010.

Kleingeld, Pauline, and Marcus Willaschek. “Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law.” *Philosophers’ Imprint* 19, no. 6 (2019).

Korsgaard, Christine M. “Constitutivism and the Virtues.” *Philosophical Explorations* 22, no. 2 (May 4, 2019): 98-116.

———. *Creating the Kingdom of Ends*. Cambridge; New York: Cambridge University Press, 1996.

———. “From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action.” In *Kant on Emotion and Value*, edited by Alix Cohen, 33-68. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan, 2014.

———. *Self-Constitution: Agency, Identity, and Integrity*. Oxford; New York: Oxford University Press, 2009.

———. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford; New York: Oxford University Press, 2008.

- . *The Sources of Normativity*. Cambridge; New York: Cambridge University Press, 1996.
- Maagt, S de. "Constructing Morality: Transcendental Arguments in Ethics." PhD diss., Utrecht University, 2017.
- O'Neill, Onora. "Autonomy, Plurality and Public Reason." In *New Essays on the History of Autonomy: A Collection Honoring J. B. Schneewind*, edited by Natalie Brender and Larry Krasnoff, 181-194. Cambridge: Cambridge University Press, 2004.
- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon, 1992.
- Railton, Peter. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 13, no. 2 (1984): 134-171.
- Rohlf, Michael. "Immanuel Kant." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2020. Metaphysics Research Lab, Stanford University, 2020.
- Rosati, Connie S. "Agency and the Open Question Argument." *Ethics* 113, no. 3 (April 1, 2003): 490-527.
- Stern, Robert A. "Transcendental Arguments." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2019. Metaphysics Research Lab, Stanford University, 2019.
- Tenenbaum, Sergio. "Formalism and Constitutivism in Kantian Practical Philosophy." *Philosophical Explorations* 22, no. 2 (May 4, 2019): 163-176.
- Timmermann, Jens. "Reversal or Retreat? Kant's Deductions of Freedom and Morality." In *Kant's Critique of Practical Reason: A Critical Guide*, edited by Andrews Reath and Jens Timmermann, 73-89. Cambridge University Press, 2010.
- Velleman, James David. *How We Get Along*. Cambridge; New York: Cambridge University Press, 2009.
- Williams, Bernard. "Deciding to Believe." In *Problems of the Self*, 136-151. Cambridge: Cambridge University Press, 1973.
- Wolf, Susan. "Moral Saints." *The Journal of Philosophy* 79, no. 8 (1982): 419-439.
- Wood, Allen W. *Kant's Ethical Thought*. Cambridge; New York: Cambridge University Press, 1999.