UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

VITOR SOMMAVILLA DE SOUZA BARROS

CONSTRUCTING OURSELVES AS MORAL AGENTS

Belo Horizonte

2019

VITOR SOMMAVILLA DE SOUZA BARROS

CONSTRUCTING OURSELVES AS MORAL AGENTS

Tese apresentada ao Programa de Pós-Graduação em Filosofia da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Doutor em Filosofia

Linha de Pesquisa: Ética

Orientadora: Prof. Dr. Telma Birchal

Belo Horizonte

2019

100
B277c
2019

Barros, Vitor Sommavilla de Souza

Constructing ourselves as moral agents [manuscrito] / Vitor Sommavilla de Souza Barros. - 2019.

142 f.

Orientadora: Telma de Souza Birchal.

Tese (doutorado) - Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas.

Inclui bibliografia.

1.Filosofia – Teses.2.Metaética - Teses. 3.Ação (Filosofia) - Teses. 4.Construtivismo (Filosofia) - Teses. I. Birchal, Telma de Souza. II.Universidade Federal de Minas Gerais. Faculdade de Filosofia e Ciências Humanas. III.Título.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

FOLHA DE APROVAÇÃO

Constructing Ourselves as Moral Agents

VITOR SOMMAVILLA DE SOUZA BARROS

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em FILOSOFIA, como requisito para obtenção do grau de Doutor em FILOSOFIA, área de concentração FILOSOFIA, linha de pesquisa Ética e Filosofia Política.

Aprovada em 18 de outubro de 2019, pela banca constituída pelos membros:

Profa. Telma de Souza Birchal - Orientador
UFMG

Prof. Alcino Eduardo Bonella
Universidade Federal de Uberlândia

Prof. Leonardo de Melo Ribeiro
UFMG

Profa. Evandro Barbosa
Universidade Federal de Pelotas

Prof. Rogério Antonio Lopes
UFMG

Belo Horizonte, 18 de outubro de 2019.

Resumo: Esta tese é uma defesa do construtivismo metaético em sua variante humiana ou relativista. No primeiro capítulo, apresento o construtivismo metaético, em diálogo com a tradição e com seus principais expoentes contemporâneos. Em resposta a críticas de autores realistas tais como Enoch e Shafer-Landau, proponho modificações ao construtivismo de Street, culminando em uma defesa qualificada do construtivismo de matriz humiana. Em seguida, no capítulo 2, analiso a obra de dois construtivistas contemporâneos, Korsgaard e Dorsey, e concluo que suas posições não encontram respaldo argumentativo. Nos capítulos seguintes, aprofundo a reflexão sobre os fundamentos da teoria construtivista e suas consequências. No capítulo 3, discuto a noção de agente subjacente ao construtivismo metaético, adentrando dois recentes debates. Primeiro, no contexto de um debate entre H. Frankfurt , G. Watson e D. Velleman, proponho que o agente seja concebido como o conjunto coerente de seus valores. A adição do termo "coerente" à proposta de Watson almeja interromper o regresso a que está sujeita a proposta original de Frankfurt. Em segundo lugar, no mesmo capítulo, procuro oferecer uma solução compatível com o construtivismo relativista para o problema da "shmagency", tal como apresentado por D. Enoch. Em seguinda, no capítulo 4, lido com conhecidos contraexemplos à teoria em vista de sua aceitação do relativismo. Em particular, analiso o contraexemplo divisado por A. Gibbard do Calígula Idealmente Coerente. Recorrendo à filosofia da linguagem e da discordância, mostro como o construtivismo não tem problemas para lidar com casos como o do Calígula, em especial se associado ao contextualismo não-indexical. Finalmente, no capítulo 5, discuto o ideal do endosso reflexivo das razões para agir por parte do próprio agente. Contra uma tradição que remonta a Sócrates e que à primeira vista encontra respaldo no construtivismo, defendo condições estritas para quando determinado agente está justificado em questionar suas razões para agir por meio de um amplo processo de autoexame.

Palavras-chave: construtivismo; relativismo; agente; reflexão; razão prática.

Abstract: This dissertation is a defence of metaethical constructivism in its Humean or relativist version. In the first chapter, I introduce constructivism in dialogue with tradition and its most prominent contemporary supporters. Responding to realist critics such as Enoch and Shafer-Landau, I suggest modifications to Street's constructivism. The upshot is a qualified defence of Humean constructivism. Next, in chapter 2, I engage with the work of two other contemporary constructivists, Korsgaard and Dorsey, and find their work wanting in argumentative support. In the subsequent chapters, I dive deeper into the foundations and consequences of the theory. In chapter 3, I reflect on the notion of agent that underlies metaethical constructivism. First, in the context of a dispute involving H. Frankfurt, G. Watson and D. Velleman, I propose that the agent be construed as the set of her coherent values, where "coherent" is meant to interrupt the regress faced by Frankfurt's original view. Secondly, I offer a solution compatible with relativist constructivism to the shmagency objection raised by D. Enoch. Then, in chapter 4 I tackle a famous counter-example, Gibbard's Ideally Coherent Caligula. Resorting to contemporary philosophy of language and disagreement, I show that constructivism deals nicely with cases such as that of the Caligula, especially if coupled with an acceptance of non-indexical contextualism. Finally, in chapter 5, I consider the ideal of reflective endorsement of practical reasons by the agent herself. Going against a tradition that goes back to Socrates and apparently resonates within constructivism, I argue for strict conditions for when an agent is justified in questioning her reasons for action via an ample process of self-examination.

Key-words: constructivism; relativism; agent; reflection; practical reason.

Acknowledgments

I thank CNPq for funding my doctoral research for four years and CAPES for funding my research visit to the Universität Wien (PDSE). At Universität Wien, I thank Professor Herlinde Pauer-Studer for sponsoring my visit to Vienna and other professors and post-graduate students involved in the Foundations of Normativity (FoNTI) project, where I found a pleasant and inviting environment to develop and discuss my ideas.

I also thank the administrative staff of the Philosophy Department and of the Post-Graduate Programme in Philosophy at UFMG, as well as the different directors the Programme had during my period as student. I thank the many different professors I had since my undergraduate days in the Philosophy Department. They certainly shaped my thought in ways I cannot fully acknowledge. I thank specially the professors from the Ethics Research Group: my supervisor Telma Birchal, Leonardo Ribeiro, Leonardo Vieira, Rogério Lopes and, more recently, Eduardo Silva.

Eduardo has been one of my main mentors since the beginning of my intellectual life in philosophy and I thank him for the great support and constant source of insight and fruitful challenge. Leonardo Ribeiro has been a great source of knowledge and inspiration for me, ever since, around 2014, I more or less unconsciously decided to dedicate myself to Anglophone moral philosophy. I haven't ceased to learn from him, both professionally and personally, since the very first reading groups and courses I took with him.

I owe special thanks to Telma, my supervisor. I often feel I don't deserve the kindness and patience with which she treats me. And she never misses an opportunity to promote and advance my work. Last but not least, she was extremely understanding and helpful when deadlines loomed over me, especially towards the end.

I equally thank the members of my dissertation committee, Leonardo Ribeiro, Rogério Lopes, Alcino Bonella and Evandro Barbosa. I realize I put them under extreme circumstances, having to read my work on very short notice. I deeply appreciate their willingness to cooperate and apologise for the inconveniences I couldn't avoid.

Many friends and colleagues were important during the several years studying philosophy that culminate with a doctoral dissertation. I will refrain from giving names, avoiding the risk of being betrayed by memory. My only mention will be to

the colleagues and friends of the Metaethics Reading Group and the Ethics Research Group at UFMG, with whom I discussed many times the ideas now presented in the form of a dissertation.

Finally, I feel enormously privileged for having as much family support as one could get. The experience of having a group of people constantly striving to make your life better is unique and makes one feel special in way I would like everyone to be able to feel. My mother: the paradigm of the caring parent. My father: my first and greatest mentor, the constant virtual interlocutor of my arguments. João: great companion. My wife: my love, my life. I wish one day I can demonstrate the extent of my gratitude to these fantastic people.

Contents

Introduction 10

1. Metaethical Constructivism and Value Constitutivism (The Standpoint and the Procedure in Metaethical Constructivism) 17

1.1. Introduction 17

1.2. Challenges to Metaethical Constructivism 18

1.3. Modifying the Standpoint Definition of Constructivism 21

1.4. The Epistemic Role of the Procedure 25

1.5. The Normativity of Instrumental Rationality 31

1.5.1. Wide or Narrow-Scope 31

1.5.2. Conceptual Necessity 34

2. Other Constructivist Theories 40

2.1. Korsgaard's Kantian Constructivism 40

2.2. Perfectionist Humean Constructivism 46

3. What is a moral agent? 52

3.1. Finding the Agent 56

3.2. Practical Justification 59

3.3. Constitutivism and the Shmagency Problem 62

3.4. Diachronic Unity of Agency 65

3.5. Some Final Worries 68

4. Constructivism, Relativism and Persistent Disagreement 70

4.1. Introduction 70

4.2. A little more on constructivism 70

4.3. Inconsistencies 72

4.4. Metalinguistic Negotiations 73

4.5. Occasional Acceptance (empirical evidence) 74

4.6. Reacting in the face of Persistent Disagreement 75

4.7. Philosophy of Language and Moral Disagreement 81

5. Setting Limits to Practical Reflection (Against Philosophy as a Way of Life) 89

5.1. Philosophy as a Way of Life: Socratic Style 89

5.2. Anti-realist constructivism 92

5.3. Limiting Practical Reflection 94

5.4. Extending Practical Reflection 97

5.5. Problems with full-analysis 100

Conclusion 103

References 106

Introduction

This dissertation is called Constructing Ourselves as Moral Agents. It is a defence of a particular variant of so-called constructivism about practical reasons. In other words, it is mainly a dissertation about what we have reasons to do and why. It is not exactly about what in particular we have reasons to do, the particular courses of action we should engage in, but rather it is an abstract, theoretical enquiry about the nature of practical reasons, what makes them reasons, what justifies them.

Let us not be misled by the title. In the first place, by 'constructing' I do not mean exactly constructing. I will not be discussing the processes by means of which we become moral agents. So this dissertation is not another chapter in the literature on The Social Construction of (Moral) Reality, since it does not engage with the discussion about how (and whether) our moral practices are socially constructed. By 'constructing' I merely wish to indicate my adherence to constructivism about practical reasons, the justifications for which will be provided at different points along the text.

Nonetheless, even within constructivism about practical reasons and in metaethics, there are interesting considerations about the processes that engendered the current state of human moral affairs. So for example, Sharon Street has offered, at different places, both evolutionary and more philosophically conjectural reasons to explain how we may come to justifiably hold values within a purely naturalistically conceived world-view (respectively Street, 2008, 2006). Though interesting and relevant, I will elude such considerations and focus instead on the structure and justification of practical reasons, leaving aside discussions about their historical explanation.

Additionally, 'ourselves' is a reference to the fact that most of us conscious human beings, if not all, participate in practices and engage in actions that might be assessed from the moral point of view. Thus, by talking about ourselves I imply that the consequences of my arguments are meant to be valid explanations of this widespread phenomenon, and not of some merely locally relevant human practice. However, as will become clear already in the first chapter, my view construes justification of practical reasons from a strictly internalist, first-personal point of

view. So, in this regard, it would be probably better to speak of 'constructing myself' or 'constructing the point of view of the agent', as what is justified for others is not *a priori* justified for myself according to this view.

Finally, though the title restricts the domain to morality, I will be discussing practical reasons or values in general. My arguments should be seen as arguments about what we have reason to do in general, not merely from the moral point of view. Of course, there is a plausible and well-known view according to which moral considerations (maybe only moral obligation) trump all other practical considerations, so that, whenever moral considerations are involved, what the agent has (all-things-considered) reason to do will coincide with what she has moral reason to do. I myself am not convinced by this claim and I am not convinced by the larger claim that moral considerations are of a different nature, but I won't discuss any of these matters here. Even if both claims are correct, there are presumably situations in which agents have to decide what to do but no moral considerations are relevant. I want to include these kinds of cases into the purview of my thesis.

For the reasons above, it is perhaps more appropriate to speak of valuing agents, or practical agents, or simply agents (as agents are, on a first pass, individuals capable of acting for reasons), instead of moral agents. Discussions about reasons and normativity have been the bread and butter of metaethicists for quite some time now. As other subfields of philosophy have started to engage with debates about reasons and normativity, on the one hand, and philosophers within metaethics have started to venture into other territories and broaden their scope beyond ethics, lots of similarities and overlaps have been discovered. As a consequence, there is a growing tendency to adopt 'metanormativity' or other similar terms in place of 'metaethics', to characterize discussions about normativity, including practical normativity. Coming from (meta)-ethics, the one subfield of philosophy where these discussions are more consolidated, I decided to stick to the traditional label at the expense of precision. For this reason, I decided to speak of metaethics, ethics, and moral agency throughout the dissertation. This strategy also helps me to pick out my interlocutors and avoid broadening too much the scope of my reflections.

Perhaps to the surprise of the reader, I will leave aside arguably the most important topic in metaethics, namely the dispute between realism and anti-realism about practical reasons or morality more specifically. Realism and anti-realism are not clearly defined terms, but the view I will end up defending here can be characterized

as anti-realist on all or almost all accounts. There is one sense of 'realist' that will be briefly discussed in chapter 2. It is the sense of 'realist' as a synonym for 'objectivist', that is, the idea that there is one single, objective truth about what all agents ought to do under similar circumstances, regardless of each agent's different values, desires, intentions, upbringing etc., but provided they all have the same evidence available. Following this view, it is possible to speak of universal values and moral obligation *tout court*. Christine Korsgaard (1996b) holds such a view and she has called her variant of it procedural realism in the past . Dorsey's "minimal core" is also a nod in this direction, although a much less ambitious one (Dorsey 2018). As we will see, it is possible to be a constructivist about practical reasons and a realist in this sense and I will be concerned with showing some pitfalls of this position during the dissertation, especially in Chapter 2.

Another, slightly more canonical way of defining realism is in reference to the mind-independence of moral reality. According to this view, there are certain moral facts or properties that exist and place demands on agents independently of the agents' mental states. In other words, there is a fact of the matter that certain things ought to be done and others ought to be avoided independently of what real people in fact value, judge, believe or desire. These facts or properties could be features of the natural world – hence, naturalist realism – or features of a non-natural, but equally real realm – thus, non-naturalist realism. Realisms in this sense come in a variety of subtly different and complex ways. I will not be engaging with any of them in particular and will not be concerned with proving their falsehood in general.[1] The main contribution I will make to the adjudication of this dispute between realists and anti-realists is provide further support to a specific anti-realist view, non-objectivist constructivism. If successful, my arguments will only make realism less appealing by making anti-realism more tenable.

There is one common argument in favour of realism that will play an important role in the dialectic of my text. It is the idea that people ordinarily use moral language as if they were realists. This is the so-called "realist surface of moral discourse". If this thesis is correct, anti-realism carries the heavy burden of having to counter the practice of otherwise apt users of moral language. For such countering, there would have to be very good reasons. Actually, my intuition is that people are

[1] Once again, the reader may wish to consult Street's own arguments against non-naturalist realism (Street 2006) and naturalist realism (Street 2008b).

generally more adamant about having a definitive, clear answer to what they ought to do, than they are to having an answer that is equal to all moral agents. That is, moral discourse demands precise answers more than it demands universally valid answers. This is an empirical matter, though, and I will postpone to chapter 4 some sketchy reflections on the empirics of the matter. I advance that my intuition is neither vindicated nor countered by the evidence adduced, but that there are reasons to at least put the thesis of the realist surface of moral discourse on hold.

***

Let us now turn to the parts of the dissertation. In the first chapter, I lay out my preferred theory of practical reasons. It is a kind of metaethical constructivism, in that it is meant to explain all practical reasons and to fall within a tradition of constructivism going back at least to John Rawls (Rawls 1980). I define, lay out and defend a version of metaethical constructivism that draws heavily on the works of Sharon Street. In fact, a significant portion of the chapter is dedicated to amendments or supplements to Street's theory, which I otherwise fully endorse. These amendments and supplements are important, though. For instance, I will take issue with Street's formal characterization of constructivism, will deepen her reflections on instrumental normativity and will provide elements for a theory of values that might avoid some of her problems. In general, however, my views remain very much aligned with her project.

Metaethical constructivism is broader than Street's approach, so in Chapter 2 I engage with some of our competitors. The chapter doesn't purport to be exhaustive, not even of the authors considered in it. But the view Street and I defend has consequences that many other constructivists balk at. So, before dealing with these arguably unwanted consequences, it is worth considering alternative approaches within the same broader paradigm. In this chapter, I examine Korsgaard's Kantian constructivism and Dorsey's Humean constructivism. Without the ambition of having decisively debunked these views, I conclude provisionally that none of them offers a plausible, philosophically sustainable alternative and that we are, thus, bound to the allegedly unwanted consequences of Street's and my constructivism.

As will come out clearly already in chapter 1, my views about the nature of practical reasons, values and agency are deeply interconnected. Therefore, my discussion about practical reasons will inevitably lead to a discussion about what agents are, particularly agents regarded as agents capable of acting for reasons and of

asserting value judgments. Chapter 3, then, is a detour from the metaethics and the philosophy of practical reasons of the previous and subsequent chapters to the troubled waters of the theory of agency.

Since the 1970's, a hierarchical model of the agent has been associated with the work of Harry Frankfurt. Roughly, the model suggests that the agent (the "person", as Frankfurt put it) is to be identified with second order desires, that is, desires – which he calls volitions – to be moved by certain first-order desires and not others. In this model, the agent's second-order desires endorse the appropriate first-order desires thus making them hers. As many – Gary Watson being first among them – have pointed out, this model faces the threat of a regress, whereby the second-order desires would have to be endorsed by third-order desires and so on. Alternatively, Watson proposed identifying the agent with her set of values. Watson's proposal has problems of its own, but in the third chapter I suggest modifications that, if successful, are able to vindicate most of his insights. I also connect the reflection about the nature of agents with questions about practical reasons and coherentist justification as discussed in previous chapters. The third chapter together with the first chapter make up the theoretical core of the dissertation.

I said above that Chapter 1's conclusions may lead to unwanted consequences and that these consequences may lead the reader to prefer other sorts of metaethical or constructivist views. Having attempted to show that at least the other constructivist views are not very promising, in Chapter 4 I turn directly to these consequences and argue that they are not all that unwanted. My preferred version of constructivism is compatible with some relativism about value. Furthermore, because it places a strong emphasis on coherence for moral justification, it is compatible with claiming that certain intuitively abhorrent views are justified, as long as they are perfectly coherent. This is the case of Allen Gibbard's famous counterexample to constructivism, the Ideally Coherent Caligula (ICC), who is perfectly consistent is his policy of maximizing the suffering of others (Gibbard 1999).

The extreme case of the ICC provides me with the perfect occasion to discuss the complexities of moral disagreement and the different justified attitudes parties can adopt in response to their interlocutors, when disagreeing about moral or related matters. Street's constructivism and mine bite the bullet and accept that the ICC is justified. This acceptance is often seen as a major weakness of the view and a motivation to seek alternatives. By pointing out the different and complex,

argumentative and non-argumentative, but nevertheless justified attitudes and strategies at the disposal of moral agents when they disagree with more or less extreme interlocutors, my goal in this chapter is to show that accepting the ICC's justification is not that problematic. In fact, I conjecture it is the best explanation of the phenomenon. If correct, my arguments should demonstrate that this is not at all a bad consequence of the view.

Also in the fourth chapter, I refer to recent debates in the philosophy of language and disagreement between so-called indexical contextualists, non-indexical contextualists and assessor-relativists. Applying these positions to moral disagreements, I take sides with non-indexical contextualists, both because their view fits more naturally with constructivism and because non-indexical contextualism makes it possible to respond to certain critics of moral relativism such as Paul Boghossian.

In the fifth and last chapter, I discuss another kind of consequence of my view about practical reasons and justification. There is an old tradition in philosophy, going back to Plato's Socrates, which sees great disvalue in a life lived without reflective examination. In the case of reasons for action, it seems like agents have to carefully examine their purported reasons before acting on them if they really want to act the way they are justified to act. This tradition, like any other, has had supporters and detractors throughout the centuries. It seems like constructivism, with the high premium it places on reflective scrutiny and first-personal coherent justification, fits squarely within that Socratic tradition. In other words, by taking sides with this loosely Socratic approach, constructivism apparently makes the task of practical justification very demanding.

In Chapter 5, I set myself the task of showing that this tradition misses something important about practical rationality and that constructivism should not team up with it. Contrary to the Socratic view, I argue that there is value in occasionally leading an unexamined life and that we should not attempt to see philosophy as the best way of life. In this chapter, I try to come up with specific conditions under which agents are justified in seeking out further reasons for their chosen course of action. Briefly stated, I argue that there are more or less precise limits to justified practical reflection and that moving beyond these limits is more likely that not to bring about suffering and frustration, without typically adding to one's justification.

Finally, a few notes about the format and structure of the dissertation. While the whole text can be read as a monographic defense of constructivism, some of the chapters are also self-standing, that is, they can be read as independent, though correlated, articles. This is particularly the case of Chapters 1, 4 and 5. Chapters 2 and 3 are primarily designed to fill in some empty spaces in the argumentation found elsewhere or to explain why I don't take different routes there. The reason why I chose this format is that I won't be so much defending one thesis as I will be defending different aspects of one and the same view. So, for each bigger aspect of the view, one self-standing chapter is designed. This structure might at times entail repetition, as when a summary of the view is repeatedly presented, but will also make for more right-to-the-point chapters. Still, I hope the dissertation is reasonably readable.

After going through the whole text, the expectation is that the reader is left with a slightly more accurate view of human reality. The point of this dissertation is both very narrow and very broad. It is narrow for it is a defence of a slightly modified view within a very precise debate on versions of constructivism about practical reasons in contemporary Anglophone philosophy. It is very broad for it means taking a stance on questions such as what are values, what are reasons for action, how should we disagree about moral matters, how much should we think and, even more broadly, how should we live. The relatively short size of the text is in part a reflection of a tendency to concision that at times verges on obscurity, for which I apologise. But it is also a reflection of the difficulty of taking a stance on many of these crucial matters. Thus, many important things were left out, some intentionally, others due to my own limitations. But I sure hope that what made its way into the text will be able to justify its existence to the careful eyes of the reader.

1. Metaethical Constructivism and Value Constitutivism (The Standpoint and the Procedure in Metaethical Constructivism)

1.1. Introduction

Traditional constructivists in Ethics and Metaethics[2] emphasize the idea of withstanding the test of a specified procedure in characterizing constructivism. However, global constructivists – i.e., those who think that all practical reasons are constructed – were allegedly shown to face insurmountable problems when attempting to account for the normativity of evaluative judgments and for the normative authority of the specified procedure, as long as they adopted that proceduralist construal of constructivism. Critics have argued that global constructivism is a theoretically untenable view.[3] Their point is not (or, maybe, not only) that constructivism is wrong, in that it yields normative principles that are false. Roughly, their point is that it is impossible to derive *any* principles for practical normativity using the tools of global constructivism, unless one already presupposes some kind of normative principle from which to build the constructive process. If true, this makes a global version of constructivism, such as the one I will be concerned with defending here, untenable. Their point is well taken and elicits a response before one moves on to discuss consequences of the view. In order to attend to it, Sharon Street has started presenting global constructivism in terms of standpoints, rather than procedures (Street 2010). As we will see, this is a move in the right direction, but falls short of dealing with the details of the critique.

In what follows, I will present and defend a version of global constructivism, while remaining sceptical about the objectivity of practical normativity. The view has been called Humean constructivism by Sharon Street. I will draw largely on her arguments, but will present original argumentation, in Sections 3 and, especially, 4. In particular, I will argue, first, that it is not the practical standpoint in general, but rather

[2] Some people dispute that Constructivism is really a different view from other views already in the metaethical market (e.g. Enoch, 2009). Others believe that Constructivism misses the target and doesn't even qualify as a metaethical view (e.g. Hussain and Shah, 2006). I will not take up any of these questions. Whether Constructivism is just a new version of some already existing position is not as important a matter as whether the view that constructivists defend – and, in particular the version of Constructivism that I will defend in the following pages – is better suited to explain practical normativity than rival theories, as I think it is.

[3] See Shafer-Landau and Enoch (Shafer-Landau 2003; Enoch 2009). They don't use these terms, but I read their critiques as having these consequences.

the specific values composing the standpoint of a given agent that are capable, together with logical and instrumental requirements, of yielding practical normativity. This means that it will be impossible to give the practical standpoint a merely formal characterization, as Street proposes (Street 2010). Second, I will argue that the procedure (understood as reflective scrutiny) plays a minimalistic but important role in normative justification: that of *finding out* what are the relevant values an agent holds and their logical and practical entailments. Its role is, therefore, epistemological, not constructive.

I won't be able to offer a detailed defence of the superiority of global constructivism over competing views. My goals are simply to show that it is a plausible view worth considering and that it withstands the attack of the critics, if modified the way I am about to suggest. This chapter is one step in that direction. Together with the next two chapters, I believe it provides significant theoretical support for the variant of constructivism that I support. Let us postpone worries about the practical consequences of the view to the last chapters of the dissertation.

1.2. Challenges to Metaethical Constructivism

The basic idea behind constructivism about practical reasons[4] seems to be the idea that there are no normative truths (in some domain) independent of some procedure that leads to them. The inclusion of "in some domain" in the parentheses is important. For certain authors believe that only some kinds of normative truths are results of a procedure (thus making the insertion of "in some domain" necessary), while others claim that all practical normativity is constructed by a specified procedure. The first authors are usually called local or restricted constructivists. Among them, we find John Rawls[5], who limits the reach of the procedure to the principles of social justice, and Thomas Scanlon[6], who focuses on the morality of what is right or wrong. These authors assume the truth of certain normative facts and from these facts, along with other non-normative facts, attempt to extract normative conclusions for the relevant domain by means of a specified procedure. On the other

[4] In this dissertation, I restrict myself to practical reasons or practical normativity, making no claims about epistemic reasons. I'll refrain from mentioning the word "practical" every time the discussion appears next.
[5] Rawls (1980, 1971).
[6] Scanlon (1998).

hand, there are the thoroughgoing or global or metaethical constructivists, who take all normative truth to be subject to the scrutiny of the specified procedure. Notable global constructivists include Christine Korsgaard[7] and Sharon Street[8].

In this text, I wish to defend global constructivism. If not for any other reason, at least because it evades the question-begging postulation of certain normative facts as materials from which to construct other normative facts. However, the characterization of global constructivism in terms of a procedure has been seriously challenged. See these two excerpts:

> So if constructivism is to avoid dignifying the arbitrary choices of idealized agents, and if it is to avoid lapsing into realism, then it must insist that these choices are exemplary because of having been formed through exceptional attentiveness to non-moral reasons. But if the reasons that are constraining the choices of the favoured agents are not moral reasons, it is hard to see why the outcomes of the initial conditions should be definitive of morality. (Shafer-Landau, 2003, p. 43)

> In order to be interestingly constructivist, the relevant constructivist procedure would have to be ineliminable in some way. But in order to be global – that is, in order to attempt to construct all normative reasons and truths – such a theory cannot help itself to any (unconstructed) normative material with which to characterize the constructivist procedure and to motivate its ineliminability. It's not clear that these two constraints can be satisfied by one and the same theory. (Enoch, 2009, p. 332)

There are at least two critiques being levelled against constructivism here. The first is that it is hard to see how someone, regardless of how rational, could extract normative reasons out of exclusively non-normative facts.[9] If the inputs of the procedure are all non-normative, why take the output to be normative? This would be possible, so it is claimed, only if we were to "dignify" the choices of some ideal agent or some procedure that would then be (the choices of the agent or the procedure) justified in counting as normative for us. But this is the second critique: what justifies dignifying the choices of some agent or accepting the outcomes of some procedure? What makes this procedure ineliminable and, more importantly, authoritative to me?

These reflections have made the life of the global constructivist a lot harder. Apparently the only kind of coherent constructivism is local constructivism, for its proponents have no problem with introducing normative facts into the procedure and

[7] Korsgaard (1996)
[8] Street (2012, 2010, 2009, 2008).
[9] This is not completely accurate of Shafer-Landau. Unlike Enoch, Shafer-Landau is discussing specifically the difficulty of extracting *moral* reasons out of non-normative facts, not normative reasons in general. But I take his and Enoch's points to have impact on general practical normativity.

also seem to have the normative resources to justify the authority of the procedure's outcomes.[10]

Reacting to the critics, particularly Enoch, Street has reformulated her presentation of global constructivism, sidestepping the importance of the procedure, and highlighting the practical standpoint instead. Her revised view is that:

> According to **thoroughgoing** or **metaethical constructivist views**, the truth of a normative claim consists in that claim's being entailed from within the practical point of view, where the practical point of view is given a *formal* characterization. (2010, p. 369)[11]

The first thing to say about Street's revised definition of constructivism is that it conflicts with the way the view is traditionally characterized, including by its most prominent defenders. So Rawls, for example, clearly gave pride of place to the procedure when he claimed that: "Rather (for constructivism), there is no such [independent moral] order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result." (1980, 568). Similarly Carla Bagnoli, in her *Stanford Encyclopedia of Philosophy* entry on metaethical constructivism, says that "metaethical constructivism is the view that insofar as there are normative truths, they are not fixed by normative facts that are independent of what rational agents would agree to under some specified conditions of choice" (2017, 1), where "specified conditions of choice" is another name for the constructivist procedure. And even Street herself, in her first paper on the topic, emphasized the importance of the idea that a normative judgment must withstanding scrutiny, an expression she borrows from Korsgaard to characterize the constructive procedure.[12]

Despite being contrary to the tradition, there are a few advantages of adopting Street's new definition. First, in contrast with a procedure, which is something apart from myself and, thus, easily seen as arbitrary and lacking authority over me, the standpoint is naturally my own standpoint. Although we will still see the complexities

---

[10] See Section 4 for discussion about this last remark.

[11] See also, "On this reading, the bumper sticker slogan of constructivism is not, as the proceduralist characterization would have it, 'no normative truth independent of procedure,' but rather 'no normative truth independent of the practical point of view'." (Street, 2010, p. 366)

[12] "According to *metaethical constructivism*, the fact that X is a reason to Y for agent A is constituted by the fact that the judgment that X is a reason to Y (for A) withstands scrutiny from the standpoint of A's other judgments about reasons" (Street, 2008, p. 223). Similar definitions can be found in Darwall, Gibbard and Railton's classic paper (Darwall, Gibbard, and Railton 1992), in Korsgaard's texts and books and in several commentators.

of how this standpoint is characterized, it is clear that the potential for estrangement in relation to the components of my standpoint is much thinner than that of a specified procedure. For the standpoint of practical deliberation or the standpoint of the valuing agent is associated to one's practical reasons or values.[13] And constructivists are typically internalists about reasons, meaning that they see reasons as having strong connections to motivation for action, under normal circumstances. As a consequence, the components or issuances of one's practical standpoint have a *prima facie* strong claim to being motivating. Another advantage of the standpoint definition is that it's becoming more widely accepted, as recent papers start to adopt it.[14] In the next sections, we'll see to what extent the standpoint definition can ground a global constructivist theory of practical reasons.[15]

1.3. Modifying the Standpoint Definition of Constructivism

Bear in mind Street's formal characterization of Metaethical Constructivism provided above. An important point to note about it is that the notion of entailment must not include any substantive normative assumption. The absence of any such normative assumption is crucial, on pain of being characterized as a realist position.[16] Although never quite clear, Street seems to understand entailment merely as logical entailment plus a requirement of instrumental rationality. Thus, "what is entailed from within the practical point of view" consists in what follows logically and instrumentally from it.

---

[13] I discuss the connection between the agent's standpoint and her values in chapter 3.

[14] See, e.g., (Dorsey, 2018, p. 593) and (Driver 2017). Admittedly, this is not so much of an important advantage. But it is not unimportant as well. Having a common understanding of concepts is certainly useful in academic philosophy.

[15] I should briefly mention an alternative approach to the definition of Constructivism. Nicholas Southwood (2018), in a recent chapter, whose purpose is to give an overview of Constructivism, has proposed that the defining feature of Constructivism is the attention to certain rules of reasoning. He finds Street's standpoint definition an improvement in comparison to the proceduralistc definition, but points to particular problems it might face and especially notes that Street's definition excludes certain views he thinks should be counted within Constructivism, such as deliberative or prudential/desire-based views. My goal here isn't to argue for the best over-arching definition of Constructivism, but to argue for the best *version* of Constructivism, which I think should give a special part to the standpoint. For that reason, Southwood's definition will not be adopted.

[16] "To explain this sense of entailment, we needn't make any substantive normative assumptions – for example, about what anyone *should* or *ought* to do or infer, or about what counts as a *normative reason* for what; to make such assumptions would be uselessly question-begging from a metaethical point of view" (Street, 2010, p. 367)

Let us postpone to the end of this chapter any discussion about Street's conception of entailment and let us provisionally accept it without criticism. It is only apparently, I claim, that this formalistic account of constructivism solves the problems presented by Enoch and Shafer-Landau in the quotations in Section 2. In particular, in this formalistic account, there are still no materials out of which to construct the normativity of further normative judgments. It is just that, instead of speaking of a procedure that derives normative truth from certain materials, we now speak of entailment. But we may ask, entailment from what?

Street equivocates between saying, more often, that normativity stems from what is entailed by the practical standpoint and saying things such as: "normative facts are constituted by facts about what is entailed by the 'rules of practical reason' in combination with the non-normative facts." (2010, 373). But "rules of practical reason", most prominently the requirement of instrumental rationality, don't have the capacity of yielding normativity, when merely combined with non-normative facts.

Some authors think practical reasoning has this capacity. In a standard, Hume-inspired approach to practical reasoning, our ends are defined by our desires or preferences, and rationality only features in the calculation of the best way of achieving those predefined ends. Contrary to that trend, some philosophers, like Korsgaard (e.g. 2011), contend that the "activity of reason" (or rationality) is capable of providing the agent with substantive reasons for action, so long as she applies her practical reasoning appropriately, that is, in accordance with what is constitutive of practical rationality. For Korsgaard, the categorical and the hypothetical imperatives are constitutive of rational agency and, by acting in accordance with them, the agent will be provided with substantive reasons for action under particular circumstances. It is beyond the purposes of this chapter to discuss Korsgaard's sophisticated view in detail. I only superficially present it to recall that Street (and I) is sceptical about Korsgaard's project.[17] But it is not like Street and I are suggesting that we stick to the standard story, where rationality has no impact whatsoever on the content of an agent's ends.[18] In my view, values are *constrained* by rationality. When an agent endorses a given evaluative judgment, this judgment has certain logical implications and endorsing it subjects the agent to a requirement of instrumental rationality. These

[17] Street's criticism of Korsgaard is discussed in almost all her papers, for instance Street (2008, 2010, 2012). In the next chapter, I engage with Korsgaard's thought and show why I don't think her arguments succeed.

[18] This is one of the reasons why I don't find the label Humean Constructivism entirely appropriate.

two facts combined make it the case that certain values are necessitated or produced (because the agent has to value them in order to continue valuing the first value) and certain other values are excluded (for endorsing them would be tantamount to not endorsing the first value).[19] So rationality does play a part in defining our ends and isn't all that inert as the standard picture suggests. However, and crucially, rationality is *not* capable, in the picture I am defending as compatible with Street's overall project, of getting an agent to value anything if that agent doesn't have any other previous values that rationally constrain her future evaluative judgments. Rationality – or "the rules of practical reason" – combined with the *capacity* to value – not any value in particular – is, thus, *not* capable of yielding any normativity in terms of practical reasons. In other words, rationality only moves us from reasons to reasons. It doesn't produce reasons out of itself plus non-normative facts.[20]

Also, when Street adopts her more common definition and claims that normativity stems from the practical standpoint, she adds that the standpoint should be given a formal characterization. The standpoint of an agent is, roughly, composed by her set of values, the set of things she takes herself to have reason to do. Valuing is, roughly, the capacity to take certain things as reasons for action. Giving the practical standpoint a merely formal characterization would include only the capacity for valuing and the rules of practical rationality such as the requirement of instrumental rationality, but would *exclude* the contingent values an agent happens to hold. But, as said before, without any values or normative judgments, there isn't anything out of which to extract further normative truths; there is no prior normative judgment that would entail anything normative. To sum up, then, giving the standpoint a formal characterization doesn't help Street avoid the difficulty with trying to come up with normative truths out of agents with no prior normative commitments.

Street's framework has powerful resources to deal with this problem. In order to do this, however, we must abandon the attempt to characterize the practical

[19] I discuss below, in Section 4, what happens in cases of conflicting values.

[20] I realize the claims in this paragraph are sketchy and obscure. At this point in the text, I am only concerned with arguing that agents must have adopted certain ends as normative for them – must have certain values – in order for it to be possible that other reasons or values are constructed out of them. That is the idea behind the claim that "rationality only moves us from reasons to reasons". Later in this chapter, I show how this "move" operates in the case of instrumental reasoning. Although I may not be able to offer a fully-fledged theory of values, remarks in that direction are offered in Chapter 3, where I discuss what it means to be a practical agent and how the notions of value, reason and agency interrelate.

standpoint in merely formalistic terms. On the contrary, we should make clear room for contingently held values in our definition of metaethical constructivism. For that reason, I proposed a revised definition as follows:

Definition: According to **thoroughgoing** or **metaethical constructivist views**, the truth of a normative claim consists in that claim's being entailed from within *a particular* practical point of view, where that practical point of view is constrained by *the constitutive features of valuing* and by *the particular values the agent holds*.

Despite Street's official definition having a formalistic character, it is likely that she would agree with the slight modification I am suggesting. In this passage she is clear:

> On this view, 'pure practical reason' – in other words, the standpoint of valuing or normative judgment as such – commits one to no specific substantive values. Instead, that substance must ultimately be supplied by the particular set of values with which one finds oneself alive as an agent – such that had one come alive with an entirely different set of evaluative attitudes, or were mere causes to bring about a radical shift in those attitudes, one's reasons would have been, or would become, entirely different.(2010, 370)[21]

The basic point I have been making, then, is that *nothing follows from valuing as such, only from particular values*. By only knowing what valuing means, you don't get substantive reasons for action. You need what valuing means (constitutive requirements of rationality and the capacity of valuing) plus specific contingent values from which to extract further normative facts (via the application to them of what is constitutive of valuing).

This revised definition is, in my view, an improvement in comparison to Street's merely formalistic definition. However, as we'll see in the next section, there is a further point that needs to be modified.

[21] See also this passage: "No one is claiming that it is entailed from within the practical point of view *as such* that *all* valuing creatures, regardless of their particular, contingent starting set of values, have normative reason to collect stamps; this would be a crazy position. Rather – I would argue – what *is* entailed from within the practical point of view as such is that a valuing creature's normative reasons depend in a certain way on *that creature's* contingent evaluative attitudes" (2010, p. 367, fn. 16).

1.4. The Epistemic Role of the Procedure

What is it we do when we evaluate our own evaluative judgment's justification? At the very least, we scrutinize the judgment, that is, we reflect to see whether it stands reflective scrutiny. In this section, I argue that it is only this kind reflective scrutiny that constructivists should mean when they highlight the importance of a procedure in establishing normative truth.

Recall that one of the main problems critics saw in constructivism, in its original, proceduralist definition, was that global constructivists seem to lack normative materials to serve as input into the procedure and, thus, ground the normativity of the procedure's output. Local constructivists don't face this challenge because they assume certain normative postulates. We saw that Street's framework, in particular after the adoption of the revised definition I proposed, is equipped to respond to this challenge. For the practical standpoint of contingent agents has abundant normative material out of which to construct: their contingently held values. However, there is another, related critique that remains without a proper reply, namely that procedures arbitrarily produce reasons. In other words, there seems to be a need for a reason to accept the procedure's authority as producer of reasons. Without further argument, there isn't yet a convincing case against an agent's feeling of estrangement in relation to any given constructivist procedure. The way I see it, this problem is insurmountable. There is no way I can rationally convince an agent to submit her current values to a given procedure if she doesn't already value (and submit herself to) the authority of the procedure.

At least, there is no way I can convince that agent to submit herself to the reasons issued by such an extraneous procedure. To deal with this problem, I believe we should reconceive the role of the procedure – reflective scrutiny – in practical justification. In particular, I argue that we should take sides with realists, who claim that the procedure is of *epistemological*, not of constructive significance.[22] This means that the procedure is conceived of as a method through which we *discover* the normative truth. The reason for this move is that if the procedure doesn't produce any new normative facts, then naturally the hard question of why submit oneself to a

[22] See (Scanlon 1998) for this kind of view.

procedure's issuance doesn't get off the ground. And this is so simply because there is nothing constructed via the procedure that claims normative authority over the agent.

Before laying out the details of this view, it's important to stress how much it contrasts with just about any other constructivist view. If there is anything like a consensus about the procedure's role among constructivists it is that it produces the correctness or truth of what it entails, rather then merely discovering its correctness or truth. Just to stick to our main interlocutor:

> In metaethical constructivism […] the fact that a normative judgment withstands scrutiny in reflective equilibrium is understood to be not only of epistemological significance but also of constitutive significance; in other words, this fact is understood to be not merely an *indication* that the normative judgment is correct, but what it *is* for that judgment to be correct. (Street, 2008, p. 238-9)[23]

This consensus, I argue, misses where the construction[24] is really happening. This is how I conceive of reflective scrutiny's epistemological role: When evaluating the justification of a given evaluative judgment (J) we happen to assert, to entertain in thought or to act on, we are *not* in the business of constructing J's normative status out of the materials of our other evaluative judgments or values (M). This enquiry into J's justifiability is rather an attempt to *find out* if J already figures in M, either in the form of a value already explicitly held, or in the form of a logical or instrumental consequence of the values already explicitly held. Reflective scrutiny – the enquiry, the procedure, practical reflection, you name it – seeks to discover if J fits together with my set of values (M). For this reason, in my view, *the idea that a judgment withstands scrutiny by being the product of a procedure* should be replaced by *the idea of a judgment that withstands scrutiny by being assessed according to the standard of one's already held values.*

I said above that the idea I am criticizing misses the point where the construction is really happening. But so far, I have only pointed to where it is not, namely in the procedure. Let me try to explain this by putting together what we have established so far. Remember the revised definition of constructivism I proposed:

---

[23] See also this passage, where she discusses Rawls: "What makes a view constructivist is its claim that the results of reasoning according to a certain procedure are correct *because they issue from that procedure*—that to be correct *just is* to issue from that procedure" (2008, p. 212).

[24] Construction or constitution or production of value. At this point, the three terms are interchangeable.

Definition: According to **thoroughgoing** or **metaethical constructivist views**, the truth of a normative claim consists in that claim's being entailed from within *a particular* practical point of view, where that practical point of view is constrained by *the constitutive features of valuing* and by *the particular values the agent holds*.

What does it mean to say that a claim is "entailed from within a particular practical point of view", what are the "constitutive features of valuing" and what role do the "particular values the agent holds" play? The "constitutive features of valuing" are the rules governing the entailment, which we are loosely defining as logical entailment and a requirement of instrumental rationality. We saw in the previous section that the "particular values the agent holds" are what the normative judgment now under scrutiny is or isn't entailed from. But what is doing this entailment? The traditional answer is: the procedure, the reflective scrutiny itself. But I believe this is an inadequate answer.

Consider. It is values themselves that have certain constitutive features, the features that preside over any instance of valuing (again, logical entailment and the requirement instrumental rationality). And it is the "particular values the agent holds" *themselves* that, because of their constitutive features as values, entail certain other values. The conclusion is that the procedure doesn't entail anything. *It's the values or normative judgments we make that, because of their nature, entail other values or normative judgments.*[25] The procedure – reflective scrutiny – has the task of finding out what are the values we hold and their entailments. But it is certain judgments (evaluative judgments, in this case) that have certain entailments. Engaging in reflection is merely a method for uncovering these entailments.[26] So whenever a normative judgment withstands scrutiny, it doesn't mean that it is *produced* by any reflective procedure. It means that the agent is *finding out* what are her values and their entailments and therefore evaluating particular judgments according to that

[25] Street doesn't seem unsympathetic to this approach, when she says that: "For one normative judgment to *withstand scrutiny* from the standpoint of other normative judgments, then, is for that judgment not to be mistaken as determined by the standards of correctness that are constitutively set by those other normative judgments in combination with the non-normative facts" (2008, p. 230). But she certainly falls short of drawing the conclusions I am drawing about the epistemological role of the procedure.

[26] Judgments have entailments. Reasoning draws inferences. In this case, reasoning draws inferences about what the entailments of a certain judgment are. I return to considerations about the nature of reasoning in the last section of this chapter, after discussing the normativity of the instrumental principle.

standard. For that reason, every assessment of a normative judgment's justification is in part a search for self-knowledge: the agent reflects and considers whether a given judgment integrates her set of values.

***

Let us take stock. In the previous section, I claimed that holding a value constitutively requires one to submit one's values to what is logically and instrumentally entailed by it. This is the constitutivist, constructive part of the story that I will explore in more detail towards the end of the chapter. In this section, I have been arguing that reflective scrutiny (the procedure), attentive to the constitute relations between evaluative judgments, finds out, discovers or brings to light my values (and their entailments). This is the epistemological, procedural, reflective part of the story. Making a normative judgment entails making many others, the ones that are constitutively necessitated by it. In the reflective search for practical justification, we bring to mind other relevant evaluative judgments we hold and we possibly discover entailments of our values that we were not aware of.

One might worry that rejecting the productive role of the procedure and favouring instead an epistemological understanding of it might amount to a rejection of constructivism. But this is not the case. Claiming that the procedure has epistemological importance doesn't mean the view is no longer constructivist. First, the traditionally constructivist resort to a procedure, although modified, has been restored. Second, I also attribute central importance to the notion of standpoint, which is gaining adepts as best overall definition of constructivism. And finally, there is still room for a constructive mechanism to operate. That is, the constitutive features of values can be rightly said to construct, i.e. necessitate, other appropriately related values. Although not all constructivists are constitutivists and not all constitutivists are constructivists, there is undeniable overlap between the two projects.

An interesting way to see how this view works in practice is to see what happens when conflict between values within a single agent emerges and one of the values has to be rejected.[27] There are two ways of understanding what goes on here.

[27] To illustrate my point, I mention cases where one of the conflicting values is indeed more deeply held by the agent. (Levy 2019) discusses a particular kind of case, in which an agent incoherently but equally values doing two conflicting things. In his example, a person is hungry and is faced with two equally valuable meals: soup and salad. He further supposes that the agent judges (A) she has all-things-considered reason to choose the soup over the salad and, simultaneously, judges (B) she has all-things-considered reason to choose the salad over the soup. As I go on to say in the main text, her incoherence makes it so that she doesn't really judge either A or B. That is, by judging A, the agent is

They might seem different, but ultimately describe the same phenomenon. One approach is to say that what will probably be rejected is the judgment that is less deeply held by the agent, since it is less representative of who the agent really is, hence of what she really values. This means that the rejected value *wasn't a real value after all*, for that agent. Of course, she could drop the other value and stick to the less deeply held. This would simply amount to a kind of self-change by her; she would probably need to make adjustments and possibly reject other values, to accommodate the new one. This change has to be more "dramatic" for her than sticking to the more deeply held value, for this is precisely what it means for a value to be deeply held: that relinquishing it amounts to abandoning a part of oneself. Another way of understanding what is going on here is to couch it in epistemological terms. Her set of moral beliefs (values) is more justified if she sticks to the more deeply held value. This is so because being deeply held means (at least in part) exhibiting a greater degree of connectedness and comprehensiveness. This means the moral belief in question is more connected to her other moral beliefs and is part of a more comprehensive set of coherent moral beliefs, namely the set of all her values, the values that constitutive her as a moral agent.[28]

Finally, notice that the view I have been arguing for explains the justification of any given practical reason an agent might be considering in terms of two basic sets of criteria. On the one hand, there is coherence and the instrumental and logical entailments of the other judgments an agent holds. So, an evaluative judgment correctly, i.e. justifiably, tracks a practical reason for a given agent if this judgment coheres with the other values she endorses. And this coherence is to be understood not merely in terms of consistency with her other consciously held values, but also taking into consideration what follows instrumentally and logically from her whole set of

making it the case that she doesn't really value B. And by judging B, the agent is making it the case that she doesn't really value A. This would mean, Levy recalls, that the agent doesn't have any reason to choose the soup or any reason to choose the salad, which apparently contradicts how we would intuitively think of this situation. The case is far-fetched and might not even have real world analogues, but the theoretical point requires an answer by the constructivist. Developing a full answer is beyond my scope here. One possibility, biting that bullet, would be to say that, yes, the agent doesn't have any reason to pursue any of the two courses of action in particular (take soup or take salad), but that the agent has reason to put herself out of that uncomfortable paralysis and thus, because she is hungry and has reason to eat, she has reason to choose randomly between the two courses of action. It is important to bear in mind that this merely sketched response is not equivalent to saying the agent's reason to eat can be transmitted to a reason to have salad and a reason to have soup, a possibility correctly rejected by Levy.

[28] Cf. Sayre-McCord, 1996, p. 166 ff., where coherence is understood as evidential consistency, connectedness and comprehensiveness. I say a little more both about the epistemic justification of values and about requirements of diachronic unity of values in Chapter 3.

values. On the other hand, there is a gain in justification the more an evaluative judgment is expressive of who the agent is, of her core values. The more strongly connected an agent is to a certain value, the more this value is representative of her and the more traumatic its abandonment would be. Thus, the less justified the agent is in abandoning it.[29]

***

There are two different processes metaethical constructivists should pay attention to when accounting for the normativity of practical reasons, or so have I argued so far in the chapter. First, there are the constitutive and necessary connections between specific evaluative judgments. Given what is constitutive of values, an agent that values a given X necessarily, that is, constitutively values Y, provided that Y is entailed by X. This entailment can be described as production or construction, hence the label constructivism. A different process happens when one reflects and puts a normative judgment to the test of reflective scrutiny. This assessment of a given normative judgment's justification *vis-à-vis* one's practical standpoint is a process of discovery. It is an epistemic enterprise consisting in the discovery that certain judgments are and certain aren't justified, from one's perspective, depending on whether they respectively cohere or don't cohere with one's set of values. This reinstates the procedure to an important role, as traditionally supported by constructivists, though a differently conceived role.

Being a global constructivist doesn't leave one without resources to explain practical normativity, because it presupposes neither that there is no normative material out of which to construct further normativity, nor that the procedure is arbitrarily normative for agents who don't antecedently accept its authority – the two main critiques I have dealt with. On the contrary, in the alternative I have been advancing in this chapter, the normativity comes from the contingently held values of a particular agent. And the procedure, for its part, merely assesses an occurrent evaluative judgment's justification by comparing it to the set of previously held values of the agent. For this reason, the procedure plays the epistemological role of bringing to the fore one's values and occasionally finding out their instrumental and logical entailments. By engaging in reflective scrutiny, we get to know ourselves

[29] I Chapter 3, I will argue that it is probably best to speak of a three-fold, rather than two-fold, justification, where coherence, agential identity and emotional attachment combine to add normative support to a practical reason.

better and discover what follows from our values in terms of reasons for acting in a given context.[30]

1.5. The Normativity of Instrumental Rationality

In this section, I briefly survey different ways of conceiving of the requirement of instrumental rationality and consider the impact these different understandings have on the project of metaethical constructivism. I show that there are a few alternatives compatible with the claims about constructivism that I have made above and that Street typically makes in her articles. After introducing a recent debate about the scope of the instrumental requirement and situating constructivism in relation to it, I will claim that it is best for constructivists to stick to some Kantian considerations originally made by Christine Korsgaard. This is what I interpret Street to be doing as well.

1.5.1. Wide or Narrow-Scope

Both Street's and my preferred definition of metaethical constructivism make clear reference to the notion of entailment or following and connect it to the requirement of instrumental rationality. In fact, one of Street's main examples of what follows or is entailed by the agent's standpoint is the means to an end. She says, for example, that "[v]aluing an end, in contrast to merely desiring it, constitutively involves valuing what one is fully aware is the necessary means to that end" (Street 2012, 44) and speaks of "what follows (as a logical and instrumental matter) from [one's] own set of evaluative commitments" (Street 2010, 371).

Strictly speaking, however, this might be inappropriate parlance. While showing the limitations in Street's definition of constructivism in terms of what follows from a standpoint, Nicholas Southwood (2018) makes a fair point. He recalls (p. 14), following Broome (1999), that the means-end coherence rule can be interpreted in two ways: narrow- or wide-scope. In the narrow-scope interpretation, if I have sufficient reason to X and believe that doing Y is a necessary means to X-ing, then I have reason to Y. On this interpretation, a reason for the means, Y, can be

[30] In Chapter 5, I discuss to what extent should an agent seek the kind of self-knowledge required for the complete assessment of an alleged practical reason's justification.

detached and, therefore, can be said to be *entailed* by the reason for the end, X, plus the requirement of instrumental rationality. On Broome's interpretation, however, this normative requirement is wide-scope, meaning that it applies to the whole conditional. In this case, what is required is that I have to see to it that (if I consider that I have sufficient reason to X and believe that Y is a necessary means to X-ing, then I consider that I have reason to Y). If this is how the instrumental requirement is to be understood, then no reason to intend the means, Y, can be detached[31] and no entailment relation obtains. In fact, as Southwood correctly observes, "wide-scope rules *never* require us to *have particular attitudes*. Rather, they simply require us to have certain *combinations* of attitudes. The attitudes that comprise a particular standpoint plus wide-scope rules never entail anything."(Southwood 2018, 1:14).

Indeed, there are a number of papers in the literature defending that instrumental rationality has wide scope. I won't be able to discuss this in detail here, but the main motivation for wide-scopers is the problem of detachment, which requires me, roughly, to intend the means I believe to be necessary for my end (not having the further options of dropping the end or stopping believing that this means is necessary for that end). Particularly problematic is the case of an absurd means that I will be required to intend, given my end and what I believe to be the necessary means to achieving that end. In Way's example, I intend to get a job and believe the necessary means to getting it is killing my rival. Narrow-scope requires me to intend the means, to kill my rival, and this seems like an unpalatable consequence of the view, to say that I have reason to or even ought to kill my rival (Way 2010).

The narrow-scope view fits nicely with Street's view. In narrow-scope terms, the appraisal of the example above would have to be one of the two following options. If I consider killing my rival absurd, that is, if I don't value it, if I disvalue it and take it to be atrocious, then the truth is that I never intended the end of getting the job *if, all along, I believed that in order to get the job I would necessarily have to kill my rival*. Getting the job simply wasn't what I valued all along, despite my belief to the contrary. This works for most people's values, I suppose. The second option is

[31] As Broome puts it: "In general, intending an end normatively requires you to intend what you believe to be a necessary means. It does not give you a reason to intend what you believe to be a necessary means." (1999, p. 410). Also here: "But instrumental reasoning does not provide you with a reason to take a means. That is not how it works. Willing (or intending) an end normatively requires you to will whatever you believe is a necessary means to the end." (1999, p. 418). Broome has since revised some of his views on this matter (see his (2013)), but his classic 1999 paper is useful to set the terms of the debate.

that I don't disvalue (that much) killing my rival, that is, I either see no problem at all with it or consider the wrongness of it as being of lesser importance than the goodness of getting the job, given my values. In this case, indeed, I value and in fact should kill my rival when this is a necessary means to getting the job. If the reader finds this second alternative counterintuitive it is probably because the reader shares my values to the point of considering killing others an inappropriate method for getting jobs. However, there is no such a thing as an obligation to hold such value.[32]

On the other hand, there is a non-negligible chance that the wide-scope view is correct and if that is the case some amendment to the theory might be due. The main salient consequence is that there will be fewer "constructed" values or normative judgments than originally thought. That is, there will be fewer values that are constitutively entailed by the mere fact that the agent holds other specific values. Instrumental rationality, unlike other constitutive relations such as analytic or logical requirements, won't have the effect of multiplying the number of normative judgments the agent (perhaps implicitly) holds. If it ranges widely, instrumental rationality will only require her either to abandon the end or endorse the means; hence, it won't necessarily be "productive" or "constructive"; it may just be eliminative of unjustified judgments (the end).

Nonetheless, metaethical constructivism shouldn't fear the narrow scope. Consider. A metaethical constructivist is most naturally an internalist about practical reasons. This means that there is no situation in which I *genuinely* take myself to have reasons for an end and in fact don't really have reasons for that end. Of course, I may be wrong about what I really take to be my reasons, that is, I may be wrong about what I actually value. In that case, reflection will help me see that I don't really have reasons for what I was erroneously taking myself to have reasons for. And in that case, of course, the "reasons" for the end won't mean I have reasons for the necessary means, simply because I didn't have (or *genuinely took myself to have*) reasons for the end in the first place. Metaethical constructivists differ on what we can *genuinely* take ourselves to have reasons for. Korsgaard's view is more demanding, but faces

[32] A further point is about the requirement of belief consistency. Once I realize I value not-p, which contradicts what I believed, namely p, then I have conflicting beliefs and, as claimed above, have to drop the belief that doesn't reflect what I value. One way of seeing what happens is an appeal to a rational requirement of belief consistency. But it's also possible to adopt Kolodny's evidentialist view (Kolodny 2005). In that case, the explanation could be that I realize I don't have a reason to believe in p. Given my values (the evidence for moral beliefs), I now know what I have reason to believe and, thus, adjust my beliefs accordingly.

problems often considered insurmountable.[33] Street's view and mine are much less demanding and, as such, don't rule out as necessarily unjustified certain values and practical reasons that we might find unpalatable.

From what was discussed so far, we can see that the narrow-scope interpretation is much more appealing to the Humean constructivist. This is so in the first place because what is instrumentally required, on this interpretation, can be said to be constructed, as the constructivist wishes to say. And, second, because Street's constructivism and mine bite the bullet that might lead someone to prefer the wide-scope interpretation. In other words, her constructivism and mine allow for an agent to have (what we might consider) bizarre reasons for action.[34]

1.5.2. Conceptual Necessity

This is one way to go. However, upon closer examination, Street seems to be making a slightly different point than simply accepting the narrow scope as stated by Broome. She appears to be reducing the instrumental to the conceptual or analytic. For example, she says:

> One cannot take oneself to have conclusive reason to Y without taking oneself to have reason to take the means to Y, where the force of the cannot here is not rational—as when one says a parent cannot rationally wish her child to be injured—but rather analytic or conceptual—as when one says that a parent cannot be childless. If someone ''judges'' that she has conclusive reason to Y, while simultaneously and in full awareness also ''judging'' that she has no reason to take what she recognizes to be the necessary means to Y, then she isn't making a mistake about what reasons she has; rather, she simply doesn't count as genuinely making the first ''normative judgment'' (or for that matter the second) at all. She's not doing what's constitutively involved in taking oneself to have a reason. (Street, 2008a, p. 228)

Her point doesn't seem so much to be that judging that one has a reason or ought to do some end entails that one has a reason to or ought to *start judging* that one has a reason for the necessary means, on pain of being instrumentally irrational. Rather, being conceptual, her point seems to be that while judging that one has reason

[33] See Bukoski (2018) for a recent critique. Street's skepticism towards Korsgaard's project is scattered throughout most of her papers. I discuss Korsgaard's work in the next chapter.
[34] I deal with a famous case of justified bizarre reasons in Chapter 4.

for an end one *is already necessarily* judging that one has a reason for the necessary means to that end, at least in so far as one is fully aware that it is a necessary means.

This reflects in her particular conception about what it means to be instrumentally irrational, a conception that contrasts with that of Broome. For Broome, the fact that you intend an end coupled with the requirement of instrumental rationality doesn't necessarily mean that you intend the necessary means. If you fail to form that intention, you are being (instrumentally) irrational.[35] In contrast, for Street, by taking yourself to have sufficient reason for end X you are simultaneously taking yourself to have reason for the necessary means Y. Crucially, if you fail to Y it is not because you fail to have reason for it. You fail because you are *not motivated enough* to do it. So, for Street, instrumental irrationality is *not* tantamount to not having the combination of attitudes required by the normative requirement of instrumental rationality. There simply is no case in which one fails to have that combination, while at the same time being aware that a given means is necessary for the end in question. The only failure that might arise happens when one fails to act as one's normative judgments dictate.[36]

Street's understanding of the instrumental principle of practical reason is largely indebted to Korsgaard. Where Street and I talk of taking something to be a reason for action or of valuing it, Korsgaard resorts to Kantian vocabulary and speaks of "willing an end" (Korsgaard 1997). She explains why her point is conceptual or analytic the following way:

> A maxim that does not already at least aspire to conform to the instrumental principle is no maxim at all. So the instrumental principle does not come in as a restriction that is applied *to* the maxim. Instead, the act of making a maxim – the basic act of will – conforms to the instrumental principle by its very nature. To will an end just is to will to cause or realize the end, hence to will to take the means to the end. This is the sense in which the principle is analytic. The instrumental principle is

[35] "If you intend to open the wine, and if you believe that in order to do so you must fetch the corkscrew, it does not follow that you intend to fetch the corkscrew. You might not have this intention if you are irrational, for instance." (Broome, 1999, p. 408).

[36] See her crucial passage on this: "A case of instrumental irrationality arises when a person is not sufficiently motivated to go ahead and do what, in virtue of taking herself to have conclusive reason to Y, she already necessarily takes herself to have reason to do—namely, to take the necessary means to Y. In this way, she fails to do what her own normative judgment says she should do. Note that normative judgments are by their very nature motivating, on my view, such that if one judges that one has reason to Y, then one is thereby necessarily at least somewhat motivated to Y. But this of course does not mean that judging that one has reason to necessarily involves a degree of motivation sufficient to result in action in every case: the opposing motivational obstacles (for instance, in the form of fear, depression, temptation, or laziness) may simply be too great." (2008a, p. 228, fn. 37)

> *constitutive* of an act of the will. If you do not follow it, you are not willing the end at all. (1997: 244)

And further ahead in the text she explains the necessary connection between willing an end and adopting the instrumental principle:

> Willing an end just is *committing* yourself to realizing the end. Willing an end, in other words, is an essentially first-personal and normative act. To will an end is to give oneself a law, hence, to govern oneself. That law is not the instrumental principle; it is some law of the form: Realize this end. That of course is equivalent to 'Take the means to this end'. So willing an end is equivalent to committing yourself, first-personally, to taking the means to that end. (1997: 245)

Contrary to merely desiring an end and contrary to simply stating (as a matter of purely descriptive, causal fact) that one will cause an end to happen, willing an end is an act of choice. It is a decision to take an end as normative to oneself and to pursue its implementation. In the language we were adopting before, taking something to be (decisive) reason to act amounts to committing oneself to its implementation and, hence, amounts to taking the means to it as normative to oneself.

This is a fruitful analysis of the requirement of instrumental rationality and one that, if correct, also allows the constructivist to justify her claim that practical normativity is constructed (entailed) when the agent follows the requirements of her practical rationality. Returning to the overall goal of this section, I conclude, then, that both Street's (Korsgaard-inspired) interpretation and the more common-sense narrow-scope interpretation of instrumental rationality can accommodate constructivism's ambitions. On the other hand, if wide-scopers are correct, then constructivism would lose an important part of its appeal, by having less normative judgments capable of being explained and justified via the constructive process. Nonetheless, for the reasons stated above, I am partial towards Korsgaard's and Street's interpretation of the instrumental principle.

***

Let me draw this chapter to a close with a few more remarks about entailment, reasoning and valuing. Reasoning is a kind of transition from attitude to attitude (McHugh and Way 2018a, 2018b), a way of attitude-revision. Entailment is a relation between propositions (including judgments and reasons[37]). How can I say at the same

[37] I am not considering here the debate whether reasons are mental states or facts.

time that reasoning is an epistemic achievement that discovers what is entailed by some propositions and say that what is a reason for the agent is what she takes to be a reason (as constructivists tend to say)? In other words, how are facts about the agent's mental attitudes and her revision of them (reasoning) and facts about the logical relations between propositions (entailment) interrelated? The answer is that what we discover in reasoning is what we already take to be a reason, but we didn't know (or, in many cases, we discover what we necessarily *would take* to be a reason for ourselves were we to entertain it in thought). But how is this different from the case of beliefs? I am also required to believe the consequences of my beliefs if I entertain them (or to drop the premise-belief). The further bite in the claim about values or practical reasons is that the agent ought not only to believe, but also to be motivated by the conclusion of the reasoning, on pain of not counting as valuing the first value (the premise) anymore.[38] This is what Korsgaard's argument about the normativity of the means being derived from the normativity of the end is supposed to have shown in the case of instrumental rationality.

Among the premises, there has to be one or more normative beliefs or values. That premise is experienced by the agent with normative force.[39] Reasoning leads to a further belief from these premises. The normative force of the conclusion is extracted from the normative force of the premise(s), not from reasoning itself. Reasoning only touches the cognitive part of values (the belief-part) and from it extracts consequences. The normative force is transmitted to the conclusion because reasoning reveals that the normative force of the premise is conditional on the normative force of the conclusion (for example, the normative force of the end is conditional on the normative force of the necessary means in instrumental reasoning). It is conditional in the sense that the agent cannot continue to take the premise (e.g. the end) as normative for her while knowingly not taking the conclusion (e.g. the necessary means) as normative for her as well.

The claim that a further value reached via reasoning will necessarily be valued by or motivate the agent, unless she drops the premise-value from which the reasoning started, I think can be made to be conceptual. If the agent resists dropping

[38] I clarify that I am talking about theoretical reasoning here, though a theoretical reasoning about practical reasons. More precisely, it is reasoning concluding in a normative belief or in what agents take themselves to have reason to do. This is not to be confused with practical reasoning in the sense of reasoning to an intention, not a belief. See Harman (1986) for this way of construing practical and theoretical reasoning (Harman 1986).

[39] I say a little more about this experience of normativity in Chapter 3.

the premise-value and doesn't come to value the conclusion, what happens is that the premise-value loses the status of a value and becomes merely a desire coupled with a non-normative belief, from which the conclusion was extracted. This is one way of showing that the desire-belief model of normative agency misses something important. The premise-belief can make the agent believe in the conclusion-belief but still fail to be motivated by it. When this happens, the premise-belief is not a real (or no longer a) value. In the kind of case I am imagining, the agent does hold a premise-belief and, associated with it, the agent is motivated to act in a way that reveals the presence of a desire. But, being a mere desire instead of a value with normative force, it will prove incapable of transmitting in reasoning its motivational force to the conclusion. What distinguishes values from desires is that values are subject to rational pressure both in their cognitive (belief) side and in their motivational side. So, values are not best defined as simply pairs of beliefs and desires.

I don't have a well defined theory of values to offer beyond the broad constrains I have presented until now and some sketchy remarks in the next chapters. In any case, one possibility is to construe values as just beliefs that are experienced in a certain qualitative, normative way, beliefs that have a certain phenomenology.[40] But if this is so, why suggest they would transmit this phenomenology to the conclusion? One possible answer is: the conclusion is part of achieving the premise, so it is really the normative force of the premise alone that is showing itself.

Let us return to the contrast between values and desires. If rational, as our discussion about instrumental rationality above has shown, the agent will value the conclusion for the conclusion is not independent of the premise-value. On the other hand, one can have a desire for something and simultaneously a desire for something else the satisfaction of which prevents the first desire from being satisfied. In other words, desires, unlike values and preferences, are compatible with contradictory ambitions. For example, if something is a necessary or better means to achieving a desire doesn't make it automatically desired by the agent (she may desire some other thing that prevents the first desire from being satisfied and, upon finding this out, still stick to desiring both).

[40] The suggestion here is that values are beliefs about what we have reason to do (or would have reason to do in appropriate circumstances) experienced in a particular (binding) way. Another suggestion is that values are beliefs about what we have reason to do to which we have a strong emotional attachment. I actually believe the truth lies in a combination of both these theses, whereby values have both phenomenological and motivating, emotional features. I say a little more about this in Chapter 3.

The idea that through reasoning we can conclude that a premise-value of ours entails some conclusion-value we didn't value or didn't know we valued seems strange. It reveals what was contained in the premise-value but we didn't know. What we claimed or judged by stating the premise before wasn't exactly what we valued, but a close relative to it. What we really value contains the conclusion (or doesn't contain the premise). This conforms to the idea that reflection is partly acquiring self-knowledge. This idea has important consequences, which I will discuss in last chapter of the dissertation. Before that, I need to get a few things clear about the theory I have been defending.

I have claimed that the normativity of the means in instrumental reasoning requires that the agent have settled an end as normative for her. The natural question that follows this claim is: how is the normativity of the end established? The view I have been defending with the help of Sharon Street leaves a lot of leeway for agents to hold conflicting values. However, Christine Korsgaard has defended a constructivist theory according to which there are strict, universal constrains on what rational agents are allowed to justifiably will. Korsgaard has put the problem very clearly we she said that "if I am to will an end, to be and to remain committed to it even in the face of desires that would distract and weaknesses that would dissuade me, it looks as if I must have something *to say to myself* about why I am doing that" (1997: 250). And for her, as is widely known, only a commitment to the categorical imperative is capable of functioning as a rational justification for the adoption of an end as normative for a rational agent. If her arguments to this effect are sound, most of Enlightenment morality will have been thoroughly and non-arbitrarily vindicated. Above, I have exposed my scepticism regarding Korsgaard's conclusions. In the next chapter, I turn to the reasons for this scepticism.

2. Other Constructivist Theories

In chapter 1, I gave an outline of my preferred version of Humean constructivism and argued for some of its premises and features. I will return to the theoretical and practical aspects of my view in the next chapters. But first I need to examine with a little more care some of the alternatives within constructivism that are available in contemporary ethics.

Ever since Rawls and Korsgaard introduced constructivism into Anglophone philosophy, in the 1980's and 1990's, there has been a virtual explosion in the literature. Thus there are several varieties of constructivism nowadays and I don't intend to analyse them all. What I will do in this chapter is briefly explore two constructivist approaches that compete with my own (and Street's). First, I will examine Korsgaard's Kantian constructivism. Next, I will dedicate myself to the analysis of an alternative version of Humean Constructivism: Dale Dorsey's perfectionist Humean constructivism. I chose these two theories because, in the first place, Korsgaard is certainly the most distinguished living constructivist, so a dissertation that doesn't minimally engage with her work and claims to be constructivist in metaethics will justifiably be looked at with suspicion. Next, I picked Dorsey's theory because it is in many ways close to my approach, since it also reclaims a Humean heritage. Finally, both Korsgaard and Dorsey propose constructivist theories that, to a smaller or larger extent, attempt to justify the existence of common reasons for action for all human beings. The reader might be feeling uneasy with the relativist notes in my constructivism. If either Korsgaard or Dorsey is right, then moral relativism should be eschewed.

2.1.Korsgaard's Kantian Constructivism

When presenting my view in the previous chapter, I said in passing that I was sceptical about Korsgaard's arguments and conclusions. In fact I don't think her arguments succeed in demonstrating what she wishes to demonstrate, namely, that all rational agents ought to conform to certain principles of rationality and that respect for these principles means respect for the core of Enlightenment morality. Now it is time to explain why Korsgaard is in my view not successful.

Korsgaard's moral philosophy is built on her interpretation and appropriation of Kant's thought. Debates over the accuracy of her Kantian interpretation are heated within Kantian scholarship. But that is not the kind of potential problem in Korsgaard's philosophy that I will be concerned with here. I will assess her arguments in their own right. From that point of view, I distinguish three kinds of potential worries for Korsgaard's approach. The first has to do with her formal conception of the rational agent. The second concerns her attempt to prove the universal value of humanity. And the third pertains to her attempt to derive the content of Enlightenment morality from the value of humanity. I will focus my critical remarks on the second worry in this chapter. Problems with her formal approach to rational agency in part mirror the problems I pointed to in Street's formal definition of constructivism. Additionally, I will further criticize her conception of the agent when I discuss what is known as the "shmagency" problem in Chapter 3. As for the third worry, it not only has already received a lot of attention in the literature, but also this worry doesn't really rise if her approach is found to be problematic already at the level of the second worry. For it is partly based on her proof of the value of humanity that she will extract the grounds of morality. Before discussing her proof, let me briefly say some general things about her proposal.

Korsgaard is a Kantian constructivist because she believes there are objective answers to moral questions, that these reasons are not real in the sense of mind-independent and that there is an appropriate procedure to arrive at these universally correct answers about morality. She is not a *robust realist* (or, as she would say, a "substantive realist") that claims there are moral facts or properties out there in the world to be *discovered* by our rational capacities. She is rather a *procedural realist*, someone who believes the true answers to moral questions are constructed by means of a procedure at the disposal of any rational agent. As long as we reason properly, argues Korsgaard, we are all capable of seeing the truth of the core of Enlightenment morality, because crucially to reason properly is to reason according to the principles of (in this case practical) rationality. And paramount among the principles of our practical rationality is what Kant called the categorical imperative.

By adopting the categorical imperative as a criterion for our reasoning, we are able to firmly ground moral obligation. In contrast with that, Korsgaard believes that substantive realism is incapable of demonstrating why we have reasons to be moral. Substantive realists claim that there are moral facts to be discovered in the world but,

once they are discovered (assuming this is possible), they are apparently left with no argument that proves that rational agents ought to comply with the moral principles they discovered. In other words, substantive realists have no answer to give to the problem of the authority of morality or the source of morality's normativity for rational agents. Contrary to that, by showing that the categorical imperative is a principle that ought to guide rational action, the challenge about the normativity of morality doesn't arise for Korsgaard. In other words, there is no question about why is it rational to do as morality dictates if following morality (in the form of being guided by the categorical imperative) just is (part of) what it means to be rational.

Thus, Korsgaard's proposal is to ground morality in our rationality. She does this with reference to the contingent sources of reasons we particular human beings have, what she calls our practical identities, that is, the social roles we inhabit and that function as ordinary sources of obligations for us. The reason why these roles are sources of obligation for us is that we, as the rational decision-makers that we are, rationally endorse them. So the reason why, say, my identities as Brazilian and as son are normative to me is that, when reflecting on what to do in particular circumstances, I endorse these roles as sources of obligation for me. Now the reason why these particular practical identities are for their part capable of functioning as reason-giving in this way is that they are, according to Korsgaard, grounded on a more fundamental and universal value: the value of humanity (Korsgaard, 1996b). We, thus, arrive at the point I want to explore in more detail, namely, Korsgaard's reappraisal of Kant's argument for the value of humanity.

Korsgaard's reads Kant's argument for the value of humanity under the supposition that "rational choice itself makes its object good", that is, that "our choices […] are value-conferring" (Korsgaard 1996a, 122) . Based on this value-conferring supposition the argument proceeds like this: if we consider that the object of our choice is good *because* it is the object of our choice, we must be committed to the idea that we are valuable ourselves, as the sources of value. Then, if we regard ourselves as valuable, we must regard humanity, our rational capacity, as valuable. And because we regard our humanity as valuable, we must regard humanity in other as valuable as well. Hence, the value of humanity is proven.[41]

[41] My presentation and criticism here, as in much ahead, is indebted to Bukoski (2018).

What is doing the argumentative work in this Korsgaardian reading of Kant is precisely that presupposition of the valuing-conferring capacity of rational agents. She argues, for example, that "[i]f we regard our actions as rational, we must regard our ends as good; if so, we accord to ourselves a power of conferring goodness on the objects of our choice, and we must accord the same power – and so the same intrinsic worth –to others." (*apud* (Langton 2007, 169). However, many have resisted this argument. Rae Langton, for example, has suggested that it is "a chain of non sequiturs" (*Ibidem*). Her point – with which many others, myself included, agree – is fundamentally that from the fact that something (A) is the source of some other valuable thing (B) or even the source of B's value, it cannot be extracted that A is itself valuable. As Langton puts it:

> We have no more antecedent reason to expect the creators of goodness to be good than to expect painters of the blue to be blue, or the creators of babies to be babies. In general we don't think the source of something valuable must itself be valuable. War can produce good poets, chicken manure can produce good roses, and in general the sources of good things can be bad. (*Ibid*.: 175)[42]

In sum, as these counterexamples reveal, the rational source of something valuable need *not* be valuable simply because it is the source of that valuable thing.[43] However, in Korsgaard's own independent work we can find additional arguments that might support the conclusion that humanity is valuable along broadly Kantian lines. That is what I turn to now.

Korsgaard's argument starts with an assumption that our rational agency has three constitutive features: a) we are reflective beings, who need to endorse one of our inclinations to bring about action; b) action is inescapable, therefore so is the problem of deciding which inclination to endorse; c) we need a justification why we choose to endorse a given inclination, a justification that is not arbitrary.[44]

Based on this conception of rational agency, Korsgaard comes up with the following argument for the value of humanity. She first holds that our contingent

[42] See also Berys Gaut's remark in the same direction: "even if we agree that the unconditional condition of an action being good is that it is the object of rational choice, it does not follow that rational choice is itself valuable. For it isn't true that if something has the power to confer some property, then the thing must possess that property. Consider the president of a university, who has the power to confer degrees. It doesn't follow that he has to have a degree himself." (1997, 174)

[43] Apparently, Korsgaard herself abandons this argument in her 1998. See Theunissen (2018) for discussion.

[44] Once more I mirror Bukoski's (2018) presentation of Korsgaard's argument and his subsequent critique.

practical identities, that is, our particular conceptions under which we find our lives worth living (or, we might say for short, our values), are the kind of thing that satisfies her third requirement above. That is, our practical identities provide us with reasons to choose not arbitrarily which inclination to endorse. However, one might question what justifies or provides value to our practical identities themselves? Korsgaard's answer is: our humanity, i.e., our capacity as human beings to confer value on objects by means of our rationally willing certain ends as normative. Her point is then that, first, we must have some practical identity or other in order to satisfy our rational need for reasons to act one way rather than another. Second, a particular practical identity only has this reason-giving capacity because something conferred it its value. For Korsgaard, the only thing that can be the source of value of any particular practical identity and about which we *cannot* further ask what gave it *its* value is our general practical identity as human beings, our humanity. In her words:

> What is not contingent is that you must be governed by some conception of your practical identity. For unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on yourself as having any reason to live and act at all. But this reason for conforming to your particular practical identities is not a reason that springs from one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as a human being" (1996b, 121)

There exists the possibility of a regress. One value (say, finishing a good chapter) is justified by another (having a good dissertation), which is justified by yet another (making a scientific contribution… or, more realistically, being able to find a good job) and so on, until, according to Korsgaard, we reach that one thing about which we cannot ask what grounds its value: our humanity. For, in her view, our humanity is what makes it possible for us to have any particular practical identity; it is a condition of possibility of our particular values and is, therefore, not only the source of their value, but also an undeniable value for rational, human beings. In other words, the value of humanity "brings a regress of justification to a satisfactory end" (1996b, 111).

There are different ways to resist Korsgaard's argument. One is to dispute that we must act in the sense that we must endorse some rational principle for our action. Maybe we are not necessarily rational, reflective agents looking for justification for

our behaviour in the way envisaged by Korsgaard. Instead of disputing her premises, I would like to point to what seems like a *circularity* in her argumentation. Bukoski has put this problem with clarity:

> Korsgaard needs to show that valuing humanity can end the reflective regress in a way that other potential justifications cannot. It would be circular to assume in the course of doing so that the value we place on other practical identities derives from the value we place (or are committed to placing) on humanity. If that claim is true, then it is true because valuing humanity can end the reflective regress, so it cannot without circularity be used in an argument to show that valuing humanity can end the reflective regress." (2018: 221)

The circularity pointed out by Bukoski is the following. Korsgaard claims both:

a) The value of humanity can stop the regress because the value of other identities derives from the value of humanity.

b) And the value of other identities derives from the value of humanity because only it can stop the regress.

Put this way, the circularity becomes obvious. Let me try to explain the point with a little more care. Korsgaard believes that, given our reflective nature, we need a justification to keep valuing a particular practical identity. But a justification to that particular practical identity can be given by appealing to any other practical (but related) identity I may have. Of course, in that case, a justificatory regress will start. Korsgaard then claims that the value of humanity alone can end this regress. The reason she gives for this is that the value of humanity is what is conferring value on all other practical identities. However, this move by Korsgaard begs the question by presupposing what one is trying to prove, because humanity is said to be what is ultimately conferring value on other identities *precisely because* it is what stops the regress.

From these considerations, I conclude this section on a sceptical note. If the circularity argument is correct, as it seems to me to be, Korsgaard has failed in two respects. First, she has failed to show why we should take what our practical identities or our values to be reason-giving, because without the grounding value of humanity we can also question our practical identities and, therefore, it seems like we are left

without a not arbitrary criterion to guide our decisions. This means that, in Korsgaard's model, we will not be able to act properly.[45]

Second, if the circularity argument is correct, Korsgaard has failed to show that humanity is valuable and not only humanity in ourselves, but also humanity in others. Therefore, her ambition to sustain the need to respect the dignity of others – the core of Enlightenment morality – on her philosophical argument for the value of humanity flounders as well. On that note, I sceptically leave my analysis of Korsgaard's moral philosophy and move on to examine Dale Dorsey's Humean alternative proposal.

2.2. Perfectionist Humean Constructivism

If the arguments presented above are correct, Korsgaard fails to offer a conception of practical reasons that can secure universal principles based on Enlightenment morality. Maybe Korsgaardians can come up with adequate answers to the challenges, but for now I assume that it is at least worth exploring alternative views within the same broadly constructivist, internalist approach.

That is what Sharon Street has set out to do and I largely agree with her Humean constructivist view, as explained in the previous chapter. Just as a reminder, in Humean constructivism an agent's reasons consist in her contingently held evaluative attitudes and their entailments. Potentially, agents may vary widely in terms of their evaluative attitudes, so there is nothing that necessarily makes it the case the all agents ought to take certain considerations as reasons for them. In other words, Humean constructivism is compatible with ample relativism about practical reasons.

Dale Dorsey has recently provided an alternative model of Humean constructivism. Unlike what he calls Garden Variety Humean Constructivism (GVHC; Street's and my view), he proposes what he calls Perfectionist Humean Constructivism (PHC). According to PHC there *are* certain basic reasons that all human agents ought to take as normative for them, regardless of their individual sets of evaluative attitudes. Dorsey motivates his view with the help of some allegedly self-evident examples. He imagines the following situation:

[45] I take up the task of showing how our values can be justified sources of reasons for action in the next chapter.

> Imagine, for instance, that we observe someone taking her child to school. I ask, "What's to be said for doing that?" You respond, "Well, of course, it's good for the child!" I respond, "Well, that can't be—she doesn't value her child's well-being." My response here just sounds off, wrong, bizarre. While it may be true that the person in question doesn't take the relevant valuing attitude, it's incredibly strange to suggest that the child's well-being counts in favor of this action in no way and to no degree at all. (Dorsey, 2018: 578)

Next, in light of the "strangeness" or "bizarreness" of evaluative attitudes such as not caring for the well being of one's child and the like, Dorsey allows himself to the conclusion that there must be a "common core" of reasons that all human beings share independently of what reasons they *take* themselves to have:

> These reflections seem to indicate that there are a subset of reasons that can be said to count in favor of our actions even if we lack the relevant valuing attitudes: the welfare of our children counts in favor of actions that advance it, for instance. This "common core" (as I shall call it) seems to include the prevention of harm, the care of the sick, beneficence rather than maleficence, and so forth. (*Ibid.*: 578)

Dorsey next considers a famous counterexample to constructivism, that of an imaginary, Ideally Coherent Caligula, whose greatest purpose in life is roughly to maximize suffering in the world. I will explore the consequences for constructivism of this imaginary figure in detail in chapter 4. For now it suffices to say, first, that being fully coherent, GVHC would apparently have to say he is justified in his life policy. Second, Dorsey is initially willing to accept –along with Street – that maybe the ambition to have strictly universal reasons is unmotivated and, thus, that maybe an altogether different rational species – like Street's rational social insects – may justifiably hold values in total conflict with our human values. Therefore, if we construct Caligula's case in a way that is so alien to our real, human lives, maybe Dorsey wouldn't object to his justified status.[46] However, Dorsey is not that liberal when it comes human values. He says: "Even if we do not wish to apply universal reasons to any alien being, say, it remains plausible to believe that real humans, in our world, should face some (perhaps small) shared set of considerations that favor action." (*Ibid.*: 580).

[46] Dorsey grants this much, but immediately after steps back from his concession: "Furthermore, and perhaps more controversially, even if we imagine an ideally coherent Caligula, would we continue to say that there is nothing to be said in favor of his refraining from cruelty? I find this hard to believe." (*Ibid.*: 580).

Dorsey's contention that there must be a minimal set of shared reasons for action at this point rests solely on the fact that he finds "bizarre", "strange" or "hard to believe" that we don't have any shared reasons. However, at least up to this point is his argumentation, there seem to be alternative explanations for why we often feel like there is a minimal core of reasons shared by all humans. Until now, I can't see why humans *should* share a set of considerations. Rather, given our shared evolutionary history and (partially) shared social lives, it is natural to suppose that humans *do* share a set of normative considerations. So, this merely descriptive fact about humans could be adduced to explain why there is (if there is) a common core to human values.

But Dorsey has a theory to explain why we should take his list of minimal reasons as *normative* to all human beings. He bases his theory on his interpretation of Hume's remarks about the standard of taste. According to Dorsey's interpretation, Hume's theory of taste is that an individual is capable of distinguishing aesthetic value to the extent that she has the appropriate sentimental reactions to the aesthetic object, where the appropriateness of a sentimental reaction is given by whether it reflects "the true sentiments of human nature" (*Ibid*.: 583). How do we know whether some instance of aesthetic appreciation and sentimental reaction mirrors what the human nature would appreciate? Dorsey's answer is: "To determine aesthetic value, there must be substantial agreement between individuals whose sentiments conform to the standard of taste; only when there is such agreement do we have evidence that human nature issues a particular verdict in a given case" (*Ibid*.: 584). So in Hume's view, according to this interpretation, the verdicts of what Dorsey calls "humanity's evaluative nature" are authoritative and whatever evaluation by an individual that conflicts with the verdicts of human nature is wrong about aesthetic value (*Ibid.*: 585).

Dorsey's plan is to apply this framework to practical normativity in general, including morality. For him, in contrast to Street's constructivism, normative content is *not* constructed from the materials given by an agent's particular evaluative attitudes. Rather, it is constructed by the evaluative attitudes of human nature in general. I would like to highlight two features of this view. The first is that, if correct, the view shows that whether an agent has reason to do something often does *not* depend on her taking herself to have that reason. And that is so not because there are real, mind-independent reasons to which agents ought to conform (as realists would

claim), but because there are reasons whose denial would not be the accepted from the point of view of human nature. That is the case, for example, of our shared reason to care for our children.[47]

The second feature is that whatever counts as the verdict of humanity is a contingent fact about the kind of being we evolved to be. As Dorsey acknowledges,

> Humanity's evaluative nature, then, is understood to be an empirical, contingent fact about the way human beings really are in our world: as a species, and as a contingent matter, we share certain values. And these shared values help to constitute our normative landscape. (*Ibid*.: 586)

Now, this acknowledgement gives rise to the kind of worry I referred to above. We might agree that there are certain values that most human beings share. But how do you conclude from that that they are *normative* to them? Especially, why conclude that they are normative to those other human beings who do *not* embrace this so-called common core (like the indifferent parent in Dorsey's example)? In other words, why take a mere regularity to be normative in a way that doesn't risk evolving into a tyranny of the majority?

Dorsey's answer to these questions is as follows:

> [H]umanity's evaluative nature is no mere statistical regularity. Humanity's evaluative nature represents, broadly speaking, the way human beings navigate the world around them and the social circumstances they face. It is a crucially important fact about human beings, for instance, that they care for their children, family, and friends—though this is certainly not a necessary feature, it is a feature of us without which our social world would be essentially unrecognizable. And this is an important advantage of PHC in explaining the normative world humans navigate" (*Ibid.*: 598)

That is how Dorsey explains why we should take the evaluative attitudes of human nature to be normative for all human beings. I find the argument unsatisfactory. In the first place, the fact that some things are very characteristic of existing humans doesn't make them, *per se*, normative. It might be that everyone holds such values, as a matter of empirical fact. Or it might be that someone who doesn't hold such values will have a really hard time living in society. In this latter

[47] As Dorsey explains in this passage: "Hence, given PHC, anyone has a reason (even if merely justifying) to care for their children. This is not because there are mind-independent reasons, but rather because her lack of a valuing attitude toward the well-being of her children will not withstand the scrutiny of humanity's evaluative nature. To fail to take the relevant valuing attitude does not indicate a failure to possess such a reason, just as the verdict of an indelicate taste is not authoritative when it comes to "catholic and universal beauty."" (*Ibid*.: 587)

case, if the agent decides to adopt the largely shared reasons of her society, the source of her compliance with human nature's evaluations is probably better understood as her desire to lead a minimally social life. Or perhaps, she values living integrated to society and believes that sharing the "common core" is a precondition for her integration. In that case, the normativity of her compliance stems from a value that she, individually, holds, not from the authority of human nature's evaluation. It is from her values that the importance, for her, of socially shared values stems.

Alternatively, one might point to the fact that humanity would probably be unrecognizably different were certain values – such as the value of taking care of one's children – to be rejected. This might be true. However, if, on the one hand, it is only a small group (or even a single individual) that rejects such values, then one can expect such rejection to have no revolutionary consequences. If, on the other, many reject it, to the point of reshaping humanity, then the case is made for these alleged core values not being part of human nature's evaluations (anymore). Finally, and independent of this last possibility, from the mere fact that a given value is considered central to the way humans navigate their world it doesn't follow that any individual within that world has any obligation to preserve or should value the preservation of such world. We are not necessarily responsible for sustaining (parts of) the moral world as it is, neither in general, nor in its current form. For all these reasons, I believe Dorsey fails to offer solid arguments from the perspective of constructivism in favour of the idea of a "common core". His view seems to be entirely founded on unjustified appeals to his intuitive responses and to merely empirical regularities.

***

I obviously didn't survey all attempts to justify a common set of practical reasons from the perspective of constructivism. My ambition in this chapter was, first, to engage with the work of the most distinguished contemporary constructivist, Christine Korsgaard. Although a lot more needs to be written before we rest our case on the merits of her approach, I believe the considerations I adduced at least justify looking for alternatives.

Given that most likely all of us don't feel confortable with the idea that someone might be justified in mistreating their child or in maximizing suffering in the world, it is natural to look for alternatives that attempt to secure at least a minimalistic set of basic reasons every human agent should have. This thought motivated my

exploration of Dale Dorsey's version of Humean constructivism. As I argued in the previous section, however, his conclusions don't seem to rest on firm ground.

For all that, it looks like we are "stuck" with Garden Variety Humean Constructivism, that is, the kind of view Sharon Street holds and that I defended in chapter 1. My goal for the next chapters is to try and turn this kind of constructivism into an appealing view, despite initial appearances to the contrary. In next chapter, I will explore some theoretical details of the view. I expect that these details will help dispel some of the doubts and resistance the reader might have entertained until now. Then, in chapter 4, I return to the counterintuitive idea that, according to constructivism, someone like the Ideally Coherent Caligula is practically justified. Hopefully, some of the reader's reservations regarding constructivism will be gone after that.

3. What is a moral agent?

This dissertation is called Constructing Ourselves as Moral Agents. I have written quite a lot so far about what I mean by 'constructing', namely, the particular version of Humean constructivism that I have been defending with the help of Sharon Street's work. In addition to that, we saw that many of the claims of metaethical constructivism about the justification of practical reasons are grounded in claims about agential unity and identity. For instance, I suggested that our values somehow make up who we are and that we typically have reason to do what reflects the values we possess and, hence, the agents we are. These claims seem a little arbitrary without further argumentation. So, in this chapter, my goal is to present a view about what it means to be an agent and how it relates to practical justification. We will see that there are a few significant presuppositions I have to make in order to make a theory of the agent play the important role I would like it to play in the justification of practical reasons. Along the way, besides presenting my favoured view, I discuss some well-known objections to it and attempt to respond to them.

At a minimal level, an agent is a being with the capacity to act, where the capacity to act is usually called agency. This is a simple but not very clear definition. In particular, it just postpones the problem from the notion of agent to those of act and agency. If it is going to be helpful to say that an agent is a being with the capacity to act, we must have a clear grasp on what it is to act, in opposition to mere behaviour. Likewise, if an agent is going to be characterized by a peculiar capacity – agency – then we must get clear on what possessing that capacity amounts to.

Another, more informative but still minimalistic, way of understanding the notion of agent is in functional terms. So, for example, List & Pettit define that an agent "is a system with these features: it has representational states, motivational states, and a capacity to process them and to act on their basis" (2011:20). States are representational if their role is to depict the world (e.g. beliefs) and motivational or intentional if their role is to motivate action (e.g. desires and intentions). Moreover, they claim that intentional states can be of different kinds:

> They may be electronic or neural configurations of the agent, for example, depending on its robotic or animal nature. They may be localized in the agent's brain or central processing system or dispersed throughout its

> body. We only require that they be configurations of the agent […] that play the appropriate functional role. (2011: 21)

List & Pettit are not particularly concerned with *human* agency. They introduce their definition of agent with the help of a fictional robot and their ultimate aim is to discuss group agency. My concern here is also not necessarily with human agency, but with a characteristic feature of human agency (maybe among other kinds of agency), namely, *conscious, rational agency*. Group agents of the kind discussed by List & Pettit are not endowed with phenomenal consciousness (List 2016). The way in which these agents process their representational and motivational states is not mediated or accompanied by any kind of distinctive phenomenology. In the famous expression, there is nothing it is like to be such an agent with their beliefs, desires, intentions or similar motivational or representational states.

In contrast with that, the idea that agents have a certain kind of *experience* of their mental states and reasons for action, that they are conscious decision-makers, is central to the claims I wish to make about justification of practical reasons. Therefore, List & Pettit's merely functional conception of agency is insufficient for my purposes. The kind of agent I have in mind is capable not only of somehow processing representational and motivational states and of acting in order to modify the world so that it resembles what the agent's motivational states posit as desirable. What I envisage is an agent that is, in Korsgaard's words, "caused to act by her recognition of certain considerations as reasons to act" (1997, 243), that is, an agent that is capable of acting based on her taking certain considerations to be practical reasons for her.

The idea of being *caused* to act by recognizing or taking some considerations as reasons to act obviously raises metaphysical worries about free causation in a deterministic world. I lean towards compatibilism, but I don't have a definite stand in this debate. For the purposes of what I am discussing, it is more important that agents are capable of *experiencing* their taking of certain considerations to be reasons for action as the causes of their actions. If metaphysically speaking agents are inefficacious and phenomenology is misleading, then our moral lives are arguably illusions, but illusions that matter to us given how we experience our mental lives.

Let me explain the point above with a little more detail. The notion of rational agency that Korsgaard and I are referring to has obvious Kantian pedigree. Kant's famous dilemma is that, on the hand, we must think of ourselves as free to be able to

have our actions guided by reason and, on the other hand, our understanding of the natural world requires that we conceive of it in a causal, deterministic way. As Dana Nelkin has put it: “Reason seems to yield the belief that one is free and that one's actions are undetermined, on the one hand, and the belief that one's actions are determined and that one is therefore not free, on the other” (2000: 564).

Kant scholars and followers have come up with different accounts to accommodate or explain away the impression that we must be committed to contradictory beliefs. Korsgaard’s own approach is to emphasize the idea of standpoints. She argues that “the standpoint from which you adopt the belief in freedom is that of the deliberating agent” (Korsgaard 1989, 38). The standpoint of the deliberating agent is also called the practical point of view, that is, the perspective from which constructivists in general explain practical normativity. From the practical standpoint, agents take considerations to be reasons and decide what to do based on those considerations.

Nelkin considers the idea of standpoints a candidate for solution to Kant’s dilemma: it accepts the contradictory beliefs, but doesn’t explain believing in the contradiction as irrational precisely because the beliefs are held “from different standpoints” (2000, 567). But Nelkin doesn’t think the proposal succeeds, because for her the idea of ‘believing from a standpoint’ is not intelligible. For her, we believe *simpliciter* (*Ibid.*: 569). I am inclined to agree with her, but even granting the possibility of holding both beliefs, she finds other problems with the standpoint solution. The main problem is that it seems difficult to appropriately sort out beliefs between the theoretical and the deliberative standpoint. In the case of the conflicting beliefs about agential freedom:

> The advocates of this two-standpoints account have the burden of explaining why it is that the theoretical belief that one is determined and not free is "irrelevant to one's deliberative task." They must say why this belief about freedom is not taken up by the deliberative point of view when other paradigmatically theoretical beliefs regularly appear to be. For one's theoretical belief that one is determined and not free would appear to be at least as relevant to one's deliberative tasks as beliefs about the effects of one's actions. (*Ibid.*: 571)

I agree with Nelkin that no adequate answer to these challenges has been given so far and I am not able to produce one myself. Hence, I am not convinced that

the two-standpoints account can explain and justify the possession by rational agents of contradictory beliefs about freedom.

Having said that, I believe the right way to go is to reject the belief that we are free *if* we are to stick to Kant's deterministic view about the natural world. In other words, *if* our theoretical reasoning takes us to the conclusion that we are not free, *then* I claim we should hold this belief from all perspectives. Whether theoretical reasoning should conclude that the world is deterministic – the metaphysical problem I alluded to above – is not something I wish to take a stand on here. What I want to argue for is that claims about rational agency do *not* necessarily have to depend on there being either metaphysical freedom, or a justified belief in freedom from the deliberative standpoint. In my view, taking considerations to be reasons for action and acting on them is possible for agents who believe they are not free. Herein lies the importance of conscious experience for rational agency.

Rational action requires that the agent *take* certain considerations to be reasons for action and that she *take* these considerations to be (part of) the causes of her action. The fact that all the metaphysical causes that in fact issued in her action, including the causes she took herself to be caused by (the reasons), may be deterministic doesn't make a difference to how she *experiences* certain causes, namely, as reasons for action. Of course, if the world is deterministic, whether or not the agent takes certain considerations to be the causes of her actions won't make a difference to whether the action will happen or not. But the important point is that the normative relation lies in the *mode of experience* of certain causes as reasons for action. Thus, it is the agent's experience, the agent's recognition, or even the agent's taking of certain considerations as reasons that characterizes the practical standpoint, not her belief in her freedom.[48]

[48] My claims in this and the preceding paragraphs are in line with the following remark by David Velleman (especially his acceptance that it might be only a matter of our self-perception, not reality) (Velleman, 2000c): "What makes us agents rather than mere subjects of behavior – in our conception of ourselves, at least, if not in reality – is our perceived capacity to interpose ourselves into the course of events in such a way that the behavioral outcome is traceable directly to us" (2000c: 128). In my view, the important thing for agency is that we perceive ourselves as acting on the basis of those causes we take to be our reasons for action.

3.1. Finding the Agent

We have characterized the agent so far in reference to what she is capable of doing, her capacity to act for reasons. But we have failed until now to identify *where* the agent is. Moral agents, conscious human beings especially, entertain a myriad of mental states and engage in countless conscious and unconscious behaviours. How do we know when the appropriate mental states are at work to turn an instance of behaviour into full-blown action? Yes, I have claimed that an agent is only an agent insofar as she takes considerations to be reasons for her and acts based on them. And I have claimed in chapter 1 that taking a consideration to be a reason means taking it to justify a course of action. But now it seems we are stuck with a circular definition that construes agents in terms of taking to be a reason and taking to be a reason in terms of agents. To make progress, we need definitions of 'agent' and 'reason' that do not immediately refer to each other. Thus, we need first to be able to pick out agents among other possible conscious beings capable of behaviour.

Which conscious beings are agents? How do we know whether an agent causes an instance of behaviour? These are questions that Harry Frankfurt has set out to answer in a series of papers since the early 1970's. Frankfurt's original concern was with sorting out cases in which the agent – which he calls 'person' – participates in the production of behaviour from cases in which she doesn't. He noticed that, unlike other animals, human beings are capable not only of first-order desires, but also of second-order desires. That is, humans not only desire that certain things happen, but also desire to have certain motives and desires. Within the class of second-order desires, Frankfurt distinguishes what he calls second-order volitions, which are second-order desires by the agent to have a certain first-order desire be her will (Frankfurt 1988) . And, as he says, "it is having second-order volitions, and not having second-order desires generally, that I regard as essential to being a person [agent]" (1988, p. 16). So, returning to our question about the identification of the agent, in Frankfurt's account the agent can be identified with her second-order volitions, with the desires she has about which first-order desire she wishes to determine her will. If the person fails to identify with any first-order desire and, thus, fails to form second-order volitions, her very condition as an agent is threatened (*Ibid*.: 21).

Influential as it may be, Frankfurt's hierarchical model has been met with criticism from early on. First among the critics was Gary Watson, who rightly questioned: "What gives these volitions any special relation to 'oneself'? It is unhelpful to answer that one makes a 'decisive commitment', where this just means that an interminable ascent to higher orders is not going to be permitted (Watson 1982). This *is* arbitrary." (1982, 108). What Watson's remarks reveal is that there is something arbitrary about saying that in particular second-order desires of a certain kind constitute the mark of the agent. In fact, our first-order desires seem to pull us in different and arguably random directions, so that the search for a second-order desires that settles the matter as to what is the will of agent (in opposition to what merely happens in her mind) seems justified. However, the same doubts we can have about first-order desires we can also have about second-order desires. For instance, why not require the endorsement of the second-order desire by a third-order desire? And so on, in a regress that is often considered to have fatally destabilized Frankfurt's proposal.

As an alternative, Watson proposed his own method for identifying conducts of proper agents. His answer is that agent's actions flow from her values or valuation system. In his definition:

> [T]he *valuation system* of an agent is that set of considerations which, when combined with his factual beliefs (and probability estimates), yields judgements of the form: the thing for me to do in these circumstances, all things considered, is *a*.
> […]
> One's evaluational system may be said to constitute one's standpoint, the point of view from which one judges the world. The important feature of one's evaluational system is that one cannot coherently dissociate oneself from it *in its entirety*. For to dissociate oneself from the ends and principles that constitute one's evaluational system is to disclaim or repudiate them, and any ends and principles so disclaimed (self-deception aside) cease to be constitutive of one's valuational system. One can dissociate oneself from one set of ends and principles only from the standpoint of another such set that one does not disclaim. In short, one cannot dissociate oneself from all normative judgements without forfeiting all standpoints and therewith one's identity as an agent" (*Ibid*.: 105-6).

Thus, Watson proposes to identify the agent whenever action stems from her valuation system, that is, her set of values. This is the kind of view I am willing to endorse, given what I have claimed so far about the standpoint of the agent and her values. Moreover, since the evaluative system that Watson writes about is not dissociable from the agent, it seems like it is not open to the kind of regress argument that plagues Frankfurt's hierarchical model.

Or maybe the agent's values *are* dissociable from him. Velleman has claimed that, just as an agent can feel alienated in relation to her second-order desires, so also in the case of values. And, so the challenge continues, these values from which the person is alienated can be the causes of her behaviour. Therefore, tracing the behaviour to an agent's values is not tantamount to tracing it to herself (Velleman 2000c, 134). But I disagree. I believe, to use Velleman's terminology, that the contribution of values to someone's behaviour *does* constitute her contribution. The key to understanding this lies in something Velleman himself suggests but discards:

> Of course, Watson refers not just to values lodged in the agent but to the agent's evaluational system; and he might argue that values are no longer integrated into that system once the agent becomes alienated from them. But in that case, Watson would simply be smuggling the concept of identification or association into his distinction between the agent's evaluational system as his other, unsystematized values. And just as Frankfurt faced the question how a volition becomes truly the agent's, Watson would face the question how a value becomes integrated into the agent's evaluational system. (*Ibid.*: 134 fn. 33)

Indeed, in order to understand how an agent's action can be traced back to her values, these values have to be integrated in the agent's total set of values, what Watson called the agent's evaluational system. What Watson didn't have and what I would like to propose is exactly an answer to Velleman's last question, namely, how does a value become integrated into the agent's evaluational system. I prefer to frame this question as a question about what are really the agent's values or, relatedly, what the agent is *justified* in valuing, given the agent she is, that is, given the particular set of values she holds. I explore the issue of justification in relation to the concept of valuing agent in the next section. For now, I just provide my answer with a promise of an explanation. My answer is that an agent is identified with the set of her *coherent* values, so that alleged values that don't cohere with the total set of values that constitutes the agent are not justified. Hence, behaviour caused by the alleged values cannot be considered action that can be traced back to the agent. By introducing reference to coherence in my answer, I hope to dispel some of the worries faced by Watson's otherwise accurate conception of the valuing agent.

3.2. Practical Justification

With this definition of valuing agent at hand, I am now in a position to discuss the nature and modes of practical justification. At various points, I have suggested that in my view normative judgements are justified if they cohere with the agent's values and that agents should respect certain principles of rationality (at least the instrumental principle) if they are to take any consideration to be a reason for her. I have also suggested that the identity of the agent, who she is, explains in part what she is justified in claiming or doing. Finally, I have suggested that a certain kind of experience of normative reasons, i.e., their experience by the agent as reasons for her, is part of what grounds the phenomenon of normativity and its authority over the agent. In fact, practical justification is in my view three-fold.

Remember that justification for constructivism is assessed from the point of view of the agent. From that perspective, the three sides to practical normativity sketched above combine and reinforce each other. Now, the phenomemological aspect is better construed as a precondition of normativity. Given how the agent experiences the end she sets herself – as normative, authoritative – this aspect accounts for the agent's motivation to follow through in normal circumstances. It is typically accompanied by an emotive response, an emotive attachment of the agent toward the action or state of affairs whose obtaining is valued by her. Without this emotive, motivational component, the evaluative claim is experienced by the agent as a mere belief that is incapable as such to motivate the agent. However, without the location of this emotional response within the context of the agent's other beliefs and values, that response risks being a mere impulse and as such risks being able only to motivate, but not to justify the agent's behaviour.

Therefore, the emotive and phenomenological precondition of normativity needs to be supplemented by other two elements accounting for practical justification. In the first place, an agent's evaluative judgment has to cohere with her set of values and respect, at least, the instrumental principle. A normative judgment has to pass this rational test because otherwise the very notion of rational agency is destabilized. As we saw, the coherence of the set of values contributes to stop a worrying regress in our philosophical search for the core of the agent. Likewise, respect for the

instrumental principle is part and parcel of the act of taking something to be a reason for action. In other words, the instrumental principle is *constitutive*[49] of valuing.

Finally, the agent's identity itself, that is, the particular set of coherent values she endorses is the ultimate source of justification for evaluative judgments. As explained in the first chapter, what an agent has reason to do is what follows from her practical point of view in conjunction with the non-normative facts of the situation. Her practical point of view just is who she is, the way she perceives the world as an open field for action. Nevertheless, this endorsement by the agent of a certain evaluative judgment should not be regarded as some kind of freestanding act of will. It is an endorsement manifested in her valuing a certain end, that is, in her being personally and emotionally attached to it. Values are beliefs or dispositions to believe to which we have a personal and emotional attachment that manifests to us with a certain, normative phenomenology.[50]

Regarding the question about the *sources of normativity*, i.e., that in virtue of which some consideration can be called normative, it is common to distinguish between externalist and internalist views. In recent work, Ruth Chang has introduced a third category, the voluntarists, which are often subsumed under the internalist heading (Chang 2009). So, her classification goes as follows:

> Normative externalists […] locate the source of normativity in a realm of external, irreducibly normative facts.
> Normative internalists, by contrast, locate the source of normativity in mental states internal to us, and in particular, in desires and dispositions to which we are for the most part passively related. A consideration has the practical normativity of a reason in virtue of its serving or furthering our procedurally constrained desires or dispositions. […]
> Finally, normative voluntarists locate the source of normativity in us, but not in our passive states. Rather, normativity has its source in something we *do*, and, in particular, in our active attitudes of willing or reflective endorsement. (2009: 244-5)

[49] More on constitutivism below.

[50] My conception of valuing closely resembles and is largely indebted to Samuel Scheffler's. See this passage: "Somewhat more precisely, it seems that valuing any X involves at least the following elements:
1. A belief that X is good or valuable or worthy,
2. A susceptibility to experience a range of context-dependent emotions regarding X,
3. A disposition to experience these emotions as being merited or appropriate,
4. A disposition to treat certain kinds of X-related considerations as reasons for action in relevant deliberative contexts." (Scheffler, 2011: 32)

The most important example of a voluntarist in contemporary philosophy is Korsgaard, even though Korsgaard herself claims to be an internalist.[51] In fact, Chang is explicit about the difficulty in clearly distinguishing between internalism and voluntarism, because, as she says, "the distinction between 'passive' and 'active' mental states is notoriously problematic" (*Ibid.*: 245 fn. 4). This imprecision at the border notwithstanding, the distinction raises the question about how my view should be classified. My impression is that it lies somewhere close to the border between internalism and voluntarism, due to its shared allegiances to Kantianism and Humeanism. In fact, the view is hybrid. It is voluntarist in that it explains part of normativity with appeal to the agent's identity and her *taking* of certain considerations to be reasons for her. But it is also internalist in that it explains or grounds this taking in part in the agent's emotional (thus, passive) attachment to the implementation of certain states of affairs under specified circumstances.

On the one hand, our values (our reasons for action) are not merely our inclinations, desires or impulses that passively take hold of us. If that were the case we would be merely caused to pursue particular ends, but not justified in pursuing them. For we would not take them to be *normative* ends for us. However, our values do have an emotional, hence passive, grounding. It is only that this emotional aspect of values is supplemented by their being partially beliefs as well. As evaluative beliefs or dispositions to believe, our values are truth-apt, are required to be coherent with each other and are more generally subject to principles of practical rationality (in particular the instrumental principle, as discussed above). Furthermore, our values are expressive of our identities as agents. As such, unlike mere impulses or passive desires, they are more stable traces of character. Based on who we are, we endorse the normative ends we take ourselves to hold.

The rational, cognitivist or belief-like aspect of values, on the one hand, and the identity-based or endorsement-grounding aspect, on the other, account for the voluntarist part of my view on the sources of normativity. Because of these voluntarist aspects, the agent can gain from time to time some critical distance in relation to her emotive responses to circumstances of action. Conversely, the emotive aspect of values guarantees that the agent is typically motivated to act in accordance

[51] But she has shown reservations about the usefulness of this terminology. See her 1997: 215 fn. 1.

with her reasons. Finally, all three aspects of values combine to account together for the justification of practical reasons, as explained above.

3.3. Constitutivism and the Shmagency Problem

Many constructivists are constitutivists about rational agency. This means they take the normative criteria for rational action to stem from the very nature of action, and not from contingent mental attitudes of particular agents, nor from mind-independent normative facts. The appeal of constitutivism is obvious in that it promises to offer standards of normativity that are universal without the need for the allegedly spooky metaphysics of realists.

So, for example, David Velleman has claimed that the constitutive aim of full-blown action is intelligibility. On this account, an action (say, taking a glass of water) will be justified if, by pursuing its specific end (say, relieving thirst), the agent will be also pursuing (not necessarily consciously) a higher-order end: the end of doing what makes sense to her (Velleman 2000a, 2009). Or in Korsgaard's view, the standards of justification for all full-blown actions are given by the principles of rationality, namely, the instrumental principle and the categorical imperative. A will that is not governed by these constitutive principles of rational agency falls short of rational action (see e.g. (Korsgaard 2009).

David Enoch has come up with a counterargument to constitutivism that has, in the eyes of many spectators, debunked the constitutivist's ambitions. It is the so-called problem of shmagency. Enoch recalls that the content of these alleged constitutive principles of agency as well as their capacity to ground all of morality are extremely contentious matters, over which the critical literature has grappled with ever since these ideas were first published. But Enoch is willing, for the sake of argument, to grant these points to Korsgaard and other constitutivists. His shmagency objection targets another potential problem for these views, namely, the worry that rational beings, when confronted with the alleged constitutive principles of action, may not feel obliged to comply with them. Just like somebody can challenge the application of the rules of chess to herself by saying that she doesn't have any reason to start playing chess in the first place, so a sceptic of constitutivism can consistently challenge, according to Enoch, that she has any reason to start "playing the game" of

action, if constitutivists insist that in order to act an agent must strive to conform to the alleged constitutive principles of action (Enoch 2006). It is worth quoting Enoch at length on this:

> And assume that our skeptic is even convinced that – miraculously – a morality and indeed the whole of practical rationality can be extracted from the aim of self-constitution. Do we have any reason to believe that now he will care about the immorality or irrationality of his actions? Why isn't he entitled 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. I am perfectly happy performing shmactions – nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution." Has Korsgaard put us in a better spot vis-à-vis this why-be-an-agent (rather than a shmagent) problem than we were vis-à-vis the why-be-moral or why-be-rational challenges with which we – or at least Korsgaard – started? Consider again the example of the house and the shoddy builder, and suppose we manage to convince him that certain standards – standards he previously did not care about and regularly failed to measure up to – are constitutive of being a house. It seems he is entitled to respond: "Very well then, I guess I am not engaging in the project of building a house but rather in the project of building a shmouse, of which these standards aren't constitutive. So what is it to me how you classify my project?" (2006: 179)

It might be possible to respond to Enoch's challenge from within the constitutivist perspective. Seen with a little more care, the challenge has two sides to it. One is the question 'why be an agent?', so that, if someone is not an agent, the normative principles constitutive of agency would not apply to her. Another is the question that assumes that a being is an agent and asks 'why take what is descriptively constitutive of my actions to be normative, that is, to provide me with reasons for action?'. Debate around both concerns has thrived in the interested community in the last few years.[52] For my part, I do not endorse constitutivism about action, but my view can be in sense characterized as constitutivism about valuing. Thus, similar kinds of worries might arise and the possibility of shmaction, or maybe shvaluing, exists. Nevertheless, the benefit of my view, as opposed to both Korsgaard's and Velleman's constitutivisms about action, is that it does *not* posit a fixed list of principles all actions (or in my case all values) ought to comply with in order to

[52] See, among others, (Bukoski, 2017; Enoch, 2006, 2011; Ferrero, 2009; Rosati, 2016; Silverstein, 2015; Tiffany, 2012; Velleman, 2000a, 2000b, 2009)

properly count as actions (values). The possibility of shmaction arises from this alleged universal standard. But, in my case the problem doesn't arise because agents can adopt *any* value they want and still the requirements of valuing will apply to them.

In the case of constitutivism about action, the constitutive principles of action are standards for the assessment of candidate behaviours. So an intentional conduct that is unintelligible from the point of view of the agent (Velleman) or that isn't guided by the categorical and the hypothetical imperatives (Korsgaard) fails to qualify as full-blown action. In the case of my constitutivism about values, on the other hand, taking some end to be normative just is valuing. Therefore, if an agent wishes to switch from value a to value b, she will not be, in any possible case, switching from value to shvalue, because in both cases she will be taking an end as normative and, if rational, pursuing its implementation. In other words, whereas in the case of shmaction it is hypothesised that an agent could be guiding her conduct by other normative principles than those constitutivists claim are constitutive of action, in the case of my theory valuing is the condition for being guided by any normative principle at all. Valuing is the taking of some consideration, principle or standard to be normative for the agent. So any purported instance of shvaluing would in fact be an instance of valuing.

To sum up this section in connection to the previous one, recall that in valuing the agent is personally disposed to endorse a given reason for action under particular circumstances. The agent is naturally free to choose to endorse a different value from the ones she currently endorses. However, she is not free to do so without cost. For a value is (in part) an emotive impulse. Our identity as the agents we are exerts a conservative pull, justifying *ceteris paribus* that we do not change to new values.

This pull has both explanatory or causal motivations and justificatory reasons. Given that we are defining valuing in part as an emotional attachment to a certain disposition to assert a particular evaluative judgment under appropriate circumstances, then it is comprehensible that we are typically *causally* indisposed to abandon a value in favour of adopting a new one. This is the explanatory reason why our identity exerts a conservative pull in us.

From the justificatory perspective, there are two considerations to be made. First, the conservative pull is required if the agent is to be minimally effective, that is,

if she is going to be an agent after all. There might be more or less diachronic stability from agent to agent, but no propensity at all to retain in the future some of my current values and intentions completely distorts our capacity to act.[53] Second, our values are typically interconnected in that larger web of values that constitutes us as the agents we individually are. Because of this interconnection, the costs of abandoning a particular value are in general high, for they typically impact other knots in the chain. Obviously, no agent forms a perfect unity. But a will that is even synchronically completely in disunity, that doesn't form an identity in terms of some kind of agential unity, is incapable of making decisions. This failed agent will not so much act as it will behave in completely unpredictable ways, ways that are not clearly the best even for herself. That is how the nature of our values and the way they interconnect to make us who we are explain away the possibility of shvaluing and favour the inescapable normativity of our values or practical reasons.

3.4. Diachronic Unity of Agency

I said above that an agent is free to choose to endorse different values from the ones she currently endorses and that this freedom doesn't come without a cost. I would like to say now a few more words about the unification in time of our valuing agency.

It is often claimed that the diachronic unity of our plans, intentions or values is a necessary precondition for rational agency. However, some authors vehemently resist this suggestion. An author at the other extreme of the spectrum is Galen Strawson. For him, there is nothing wrong with being an "episodic self", that is, someone who doesn't see herself in the present as the same human being she was in the past and the human being she expects to be in the future (See for example Strawson (2004)).

Now I believe some kind of experienced diachronic unity is required, otherwise we can't make sense of our future plans and won't feel rationally pressed to start their implementation in the present. For unless I conceive of my future self as myself I won't experience the carrying out of my plans in the future as normative for me in the present (they will be normative to another agent, my future self). And unless

[53] More on the need for diachronic unity ahead in the chapter.

I believe the plans I devise at the present will continue to be normative to me in the future, I won't be able to make sense of anything normative whose implementation is temporally extended. For I will not be allowed to expect that it will remain normative to the future agent I will become. Thus, I can't make sense of the notion of planning if I can't see myself as the same self that I will become in the future.[54]

Michael Bratman defends the importance of certain kinds of temporally-extended unification that allow me to draw connections to the conception of valuing agent I defended above with the help of Watson's remarks. In the context of the debate between Frankfurt, Watson and Velleman that I discussed above concerning the appropriate way for an agent to identify with her first-order desires and treat them as reasons for rational action, Bratman contends (Bratman 1999):

> The key, I think, is to notice that a decision to treat [a desire] as reason-giving might itself be incompatible with the agent's *other* standing decisions or policies concerning what to treat as reason-giving. The grudging addict might have a general policy against treating his desire for the drug as reason-giving, and yet, in the face of the present urgency of the desire, he might decide to treat it as reason-giving this time. It seems to me that such an addict does not identify with his desire for the drug, even though he decides to treat it as reason-giving this time. (1999: 200)

Bratman's fragment supports the idea that plans, policies or decisions made in the present are normative for the future agent, but the mode of transmission of that normativity is not exactly that of assuming that past decisions have normative weight on the agent's future deliberation. What are doing the argumentative work in the quoted example are the *other* reason-giving considerations (decisions, policies or, as I would prefer to say, values) that the agent *presently* holds alongside the desire to take

[54] These considerations are supported by the views of several philosophers. For example, Korsgaard claims that "the choice of any action, no matter how trivial, takes you some way into the future. And to the extent that you regulate your choices by identifying yourself as the one who is implementing something like a particular plan of life, you need to identity with your future self in order to be *what you are even now*. When the person is viewed as an agent, no clear content can be given to the idea of a merely present self." (1996b: 114) And Michael Bratman, perhaps the philosopher most renowned for defending the diachronic unity of agency, argues that "it is a deep and important feature of our agency that it is temporally extended: one and the same agent persists over time, and there are complex continuities and connections that help constitute the organized interweave of our action and practical thinking over time. Indeed, on a broadly Lockean approach to personal identity, the connections and continuities that are the back-bone of this psychological, cross-temporal quilt are constitutive of the identity of the agent over time, an identity that is presupposed in much of our practical thinking. And this suggests the conjecture that it is primarily its role in constituting and supporting this organized, cross-temporal, Lockean interweave of action and practical thinking that confers on a structure of attitudes a claim to speak for the agent—a claim to agential authority." (Bratman, 2007a: 5)

the drug. As I argued for when discussing Watson's model, a consideration or desire that doesn't cohere with the agent's set of values (that is "incompatible with the *other* standing decisions or policies", in Bratman's words) fails to qualify as a reason for action.

What the point above suggests is that the normative force of past decisions or plans depends on their *presently* featuring in the agent's set of values. In other words, it depends on the agent presently taking them to be reasons for action. What explains that the drug addict in Bratman's example doesn't take the desire to take the drug to be reason-giving is that it doesn't cohere with his present set of values.

If this argument is sound, the case for a diachronic requirement of agential unity seems seriously weakened. The argument shows that agents should *not* take their past values, reasons, plans, intentions or decisions to be normative for them in the present. Or to be more precise agents shouldn't take their past values to be normative *merely because they are the values of their past selves*. As a matter of fact, agents should very often take the values they held in the past to be normative for them in the present. But that is so because, in fact, agents ordinarily hold sets of values that are largely stable in time, so that past values will most likely remain present, thus normative, values.

What these considerations show is not that the diachronic unity of agency is illusory. What they reveal is the existence of an asymmetry in how agents relate to their past and their future reasons. On the one hand, identifying with our future selves and their reasons is necessary for us to engage in temporally-extended activities in the present. On the other, identifying with the past is not necessary for rational action and is merely a usual feature of human agency. Obviously, if the agent starts to manifest a tendency towards abandoning in the present every single or almost all of her past commitments, then her capacity to believe in the present that she will follow through with the plans she currently comes up with will be severely impaired. Therefore, a modicum of unification with the past is also a prerequisite for temporally-extended rational action. But the reason for, and the extent of, this identification with the past are of a different and much less demanding nature.[55]

[55] It is actually possible for agents to completely break with their past selves and evaluational worlds. This allegedly happens in very extreme, and rare cases such as religious conversions or entering the military career. In these situations, what sociologists Peter Berger and Thomas Luckmann called "alternation" happens (Berger and Luckmann, 1966: 157-61).

3.5. Some Final Worries

Let us return to the question about the definition of agent. As I have previously defined it, the agent is a kind of locus of experienced decision making. Alternatively, we can say that the agent is the self from the point of view of her reasoned and conscious capacity to act. I said a precondition for rational agency is that the agent experiences certain considerations as reasons, that she has values. And I also said the agent can be identified with her set of coherent values, of which she has a particular kind of phenomenology.

So the agent can consciously experience the values that compose her. Can she also experience herself as a unity over and above her particular values? For example, it is often thought that the self is at least in part adequately conceived as a unified mental entity, an 'I', with which we as rational beings identity over and above our particular physical and psychological traits (G. Strawson 2017). Humeans, however, resist the idea of an identical self that accompanies all our mental states and of which we can be consciously aware.[56]

How does this dispute impact my previous arguments about the justification of rational action? For example, I claimed that agential unity is important for the capacity to act. Does it mean that the agent is required to experience herself as a (mental) unity that is efficaciously causing her behaviour for the reasons she takes to have? The answer is no. If there is such a thing as a mental self or agent of which I am conscious when I act, that thing of which I am aware just is the unity of my self's practical reasons or values. But if we are not capable of experiencing such unified agent, there is no problem. I don't have to experience myself as a unified entity for there to be a unified agent (the set of my values) that I practically am. If that is the

[56] In Hume's famous words: "For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*, and may truly be said not to exist. And were all my perceptions remov'd by death, and cou'd I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I shou'd be entirely annihilated, nor do I conceive what is farther requisite to make me a perfect non-entity. If any one upon serious and unprejudic'd reflection, thinks he has a different notion of *himself*, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continu'd, which he calls *himself*; tho' I am certain there is no such principle in me.
But setting aside some metaphysicians of this kind, I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement." (*Treatise*, 1.4.6.3-4, 2011)

case, what I experience just are the values themselves, their interrelations and the objects in the world to which they are related, but I don't experience a unity over and above them. Both views are compatible with my arguments, so I don't have to take a stand on this intricate debate.[57]

There is another problem in the vicinity. I said the agent is the locus of experienced decision making. The agent experiences certain considerations as reasons for action. And I further said that the agent is identical to her set of coherent values. But, strictly speaking, values are something an agent *has*. How can the agent *be* the values or experiences she has? This a version of a general problem for views that dangerously approximate the idea of a self or a subject of experiences with the experiences themselves: there must be some subject that has these experiences.[58]

Avoiding any commitment about the metaphysics of the mind and the explanation of consciousness, I believe the safest way to proceed is to claim that my self is whatever material basis is required for conscious experiences coupled with the particular experiences and mental states that I experience as being mine. This is true of my desires, intentions and beliefs and is also true of my future plans, as discussed above. Furthermore, while the material basis is necessary, its particular components are potentially all disposable. So the suitable material basis is defined as "that one that sustains my experiences". In principle, this allows for decay in time (slow or sudden), survival in another body or material basis (including not-human, computer generated) and even duplication.[59] In sum, the point is that the agent I am exists as long as there is a material basis that sustains (that has) the normative and other relevant non-normative experiences I have. And with these remarks, I conclude my discussion about the nature of evaluative agency that underlies several of my claims about practical justification.

[57] For further discussion on the issue of agential self-awareness, see (Bayne and Pacherie 2007).
[58] See Paul Snowdon's 2018 review of Strawson's 2017 book for an argument along these lines (Snowdon 2018).
[59] I do not claim to have settled the dispute about these famous thought experiments and counterexamples that populate the literature on personal identity. I only hint at how they could be tackled from the perspective of what I discussed previously in the dissertation.

4. Constructivism, Relativism and Persistent Disagreement

4.1.Introduction

Constructivism in metaethics typically counts consistency[60] among an agent's values or evaluative judgments as one of, if not the most important criterion for justification. However, the critical literature is fraught with imaginary examples of fully consistent individuals, who nevertheless hold intuitively abhorrent values. The most famous of them is Gibbard's Ideally Coherent Caligula (ICC). Roughly, the ICC's only policy in life is to maximize the suffering of others. Gibbard (1999: 145) And because he is perfectly coherent, it seems like constructivists are left without the tools to criticize him.

Some constructivists (e.g. Korsgaard, 1996) believe the formal features defining what counts as normative truth for constructivism are enough to rule out the possibility of the ICC. Others bite the bullet and accept the possibility of a justified ICC (Street, 2009). However, Street's acceptance of the ICC is usually seen as a confession of weakness and a reason to move away from (or at least modify) her version of constructivism (e.g. Dorsey, 2018; Schafer, 2014). In the present chapter I wish to restore the reputation of the relativistic kind of constructivism expounded by Street. I will argue we should accept the possibility of a justified ICC and that there is a host of alternatives to what can be done as part of our disagreement with him, alternatives that – I claim – can accommodate our sense of dissatisfaction with his justified status.

4.2. A little more on constructivism

Traditional definitions of metaethical constructivism stress that normative judgments only rise to the level of truth after having stood the test of some kind of "procedure" (Rawls, 1980), "scrutiny" (Korsgaard, 1996) or "specified conditions of choice" (Bagnoli, 2017). For reasons that extend beyond the scope of this chapter[61],

[60] I will use consistency and coherence as interchangeable throughout this chapter.
[61] See Chapter 1 for discussion.

there has been a recent tendency, led by Sharon Street, to define constructivism not anymore in terms of "procedures", but in terms of "standpoints".[62]

According to constructivism in the standpoint definition, a normative claim is true if it is entailed by the practical standpoint or point of view of the agent (Street, 2010: 369). It is unclear what exactly is meant by "entailed", but it is safe to assume that logical entailment and instrumental rationality are part of it. The fact that a judgment's justification is assessed *from the point of view of the agent* is important for what we will discuss later. We can call this the internalist aspect of the view.

Another noteworthy feature of constructivism is the importance given to coherence. A normative claim is true if it is entailed from the agent's total set of values and it must naturally be consistent with it. Consistency is precisely the feature of constructivism causing it the problem we are considering, i.e., the counterexample of the ICC.

Among constructivists in metaethics, there are those who believe the kinds of features described either in the procedure definition or in the standpoint definition are enough to guarantee that all rational agents should endorse an objective list of normative claims. For these constructivists – Korsgaard (1996) being the most prominent – someone like the ICC is impossible, because the formal features of constructivism prevent agents from justifiably endorsing a value such as the maximization of human suffering. For these authors, responding to Gibbard's challenge means showing that it does not really take off the ground.

The critical literature around Korsgaard's version of constructivism is very rich and I will not attempt to discuss it.[63] For the purposes of this chapter, I will assume that her solution fails and, therefore, that the ICC is possible within a constructivist framework. In fact I will assume what might be called a relativistic or subjectivist version of metaethical constructivism, according to which it is perfectly possible for different agents to be justified in holding conflicting values. This does not mean that just about any normative claim is permitted. There is a complex story to be told about how justified normative claims emerge, a story that presumably involves evolutionary[64], rational[65] and societal aspects. Moreover, given the extent to which we

[62] See Street, (2010; 2012) and Dorsey (2018).
[63] I critically discussed Korsgaard's constructivism in chapter 2.
[64] See Street (2006) for a plausible story about the role of evolution in conforming some of our normative claims.

share a natural history as well as a life in society, it is expected that most of us will have a great deal of overlap in our sets of justified normative claims.

Some philosophers are sympathetic to constructivism and share my scepticism about Korsgaard's views, but are not willing to take the relativistic step I just took. Street (2009) bites the bullet and takes the step, but most constructivists balk at the idea. Often they mention the very case of the ICC as a reason not to take this step and instead look for some kind of "minimal core" or "rational convergence" that prevents ICC's from being on a par with us.[66] My goal in this chapter is to show that Street is right in biting that bullet, by showing that the pill is not that bitter to swallow after all. It turns out there is a lot that can be said about our reaction to the ICC that is compatible with a relativistic version of constructivism.

4.3. Inconsistencies

The first thing we need to be aware of when trying to explain our disagreement with the ICC is that, *ex hypothesi*, the ICC has no problems with consistency. This is a very strong feature of the example as it is virtually impossible to come across any real person who is completely coherent in her values and judgments. It is important to bear this in mind, because it reveals both an advantage and a disadvantage of the example. The advantage is that, being able to assume Caligula's full consistency, we can focus on everything else that might be at stake. In other words, the highly stylised case allows us to focus on what we want to investigate without potential confounding effects.

But there is a downside to this. It seems safe to assume that one of the main things we do when we disagree with someone is check whether they are consistent in their claims about the subject matter. Especially in cases of disagreement about matters of value, where decisive facts to settle the matter are presumably scarce, two inconsistent statements by an opponent constitute an obvious fact to be pointed to and used to debunk them. So, if checking for inconsistencies plays such a significant role

[65] See Street (2008) for elements of a conjectural story about how asserting certain normative judgments rationally constructs and demands the assertion of other specific judgments.

[66] See Dorsey (2018) for a constructivism with a "minimal core" and see Schafer (2014) for an expectation of rational convergence. In the case of Shafer, the ICC is also part of what is motiving him to reject non-indexical contextualism and endorse assessor relativism, because the latter does not require the assessor to yield to the assessed subject's standard. I will discuss these topics towards the end of this chapter. I critically discussed Dorsey's Humean constructivism in the previous chapter.

in value disagreements, it seems reasonable to further assume that an important part of our reaction to the ICC is due to our expectation of his inconsistency and that, in removing that, some of our dissatisfaction would fade away. An empirical test would have to be run to assess how much people insist on claiming that someone like Caligula is wrong or his values are false after they find out that he is perfectly coherent. I do not claim to know the answer to that. I am just hypothesizing that some of the dissatisfaction will be gone, but very unlikely all of it.[67]

Checking for inconsistencies is the paradigmatic constructivist strategy. If we are deprived of that, what else can we do in reaction to the ICC? In the next sections I provide some insights based on philosophical and empirical literature on moral disagreements and practices of attribution of blame and responsibility.

4.4. Metalinguistic Negotiations

Another way of explaining people's reaction to the ICC is to understand it as a form of metalinguistic negotiation. A metalinguistic negotiation happens when parties dispute about the appropriate word or expression to use in a given context where a normative question is at stake.[68] These verbal disputes are often tacit, but, according to proponents of this approach, reflect a genuine sort of disagreement. In other words, we are not entitled to assuming that parties in a genuine disagreement share the meaning of the relevant words they use. Rather, they often (tacitly) advocate for their preferred usage of the words.

Could we be engaged in some kind of metalinguistic negotiation with the ICC when we react to his intuitively abhorrent, but consistent normative judgments? Well, not unless we are taking part in some kind of conversation with him. To the extent that he, on the one hand, merely holds or acts upon a given value we find repulsive and we, on the other, merely judge that value to be repulsive on our own, there seems to be no way in which we could be negotiating anything with him. More precisely, our disagreement would be only a disagreement in state, not a disagreement in activity (see Cappelen and Hawthorne (2009: 60) for this influential distinction). Each

[67] Later in the chapter, I refer to some studies in experimental philosophy about moral disagreements and relativism.
[68] See Plunkett and Sundell (2013) for a detailed analysis and defense of the view. See Chalmers (2011) on verbal disputes for a source of part of their inspiration.

party in the disagreement would not be asserting anything in the presence of the other party that the latter rejects.

But it is possible to come up with a hypothetical dialogue between a representative of some standard Western morality and the ICC. In that scenario, we could have the ICC say something like:

1) *It is right to inflict as much suffering on others as possible.*

To which the Westerner might respond asserting the denial of 1:

2) *No. It is not right to inflict as much suffering on others as possible.*

In uttering 1 and 2, the natural interpretation is to say that the Western and the ICC are engaged in a disagreement where sentences with exclusionary content are asserted. That is, 1 and 2 cannot be both justified according to the same standard. However, it might be that another, more tacit kind of genuine disagreement is taking place. Maybe parties are negotiating metalinguistically about the appropriate use of the word "suffering". So maybe, when the Westerner disagrees with the ICC, one of the things she is trying to do is, say, try to make the word "suffering" be associated only with acts that are *not* right.[69]

These cases of disagreement not directly related to the content expressed in the opposing assertions are plausibly common and the recognition of their existence has been an importance addition to the literature. Nevertheless, it seems very inadequate to assume that all possible disagreements we might have with an Ideally Coherent Caligula can be reduced to instances of metalinguistic negotiation and similar phenomena. These phenomena account for some of the data and help us give a more complex picture of disagreements. But that picture needs to be further completed.

4.5. Occasional Acceptance (empirical evidence)

Recent empirical studies have been testing people's intuitions about relativism and disagreement. Some of them have suggested that participants often reject the claim that at least one of the conflicting judgments in cases of moral disagreement must be incorrect (Sarkissian et al. 2011; Beebe 2014; Khoo and Knobe 2016).

[69] There are other kinds of disagreement without exclusionary content discussed in the literature. See, for instance, Khoo and Knobe (2016) for an explanation of disagreement as an attempt to produce change in the conversational context.

According to the last of these studies, the more culturally distant the disagreeing parties are, the more participants are willing to accept that neither of the conflicting judgments must be incorrect.

If these studies are correct, maybe the ICC is not a real threat to constructivism after all. Critics might be relying on their own prejudices – that at least one of the parties to a disagreement must be wrong – and confusing their prejudices with a widely shared view among ordinary users of moral discourse. Contrary to that, these studies suggest that people do *not* presuppose that at least one of parties to a moral disagreement is wrong when the disagreement is among culturally far away individuals. Therefore, if the studies are correct, most people might *not* find the ICC to be wrong (under the natural supposition that the ICC is someone culturally far away from these evaluating individuals). Therefore, a view – like the version of constructivism that I favour – that accepts the ICC's justification would not be countering the intuitions of ordinary users of moral language.

This might be so. Still, even if the studies are proven right, there is still a big minority of participants claiming that the distant parties to a disagreement cannot both be right. So, there is at least a big minority of cases of resistance to accepting the justification of the ICC and other similar characters.

The following sections attempt to come with an explanation for why is there such resistance and also for why the resistance seems to reduce the farther away the disagreeing parties are from each other. Just like metalinguistic negotiation, it seems that simple acceptance is part of the range of options at the disposal of disagreeing parties when faced with persistent disagreement with radical figures. However, there are still other alternative explanations to which the constructivist can appeal to account for the occasional remaining resistance. Furthermore, these remaining strategies help us understand why the resistance fades the more culturally far away parties to a disagreement are from each other.

4.6. Reacting in the face of Persistent Disagreement

So far we have seen that constructivism typically explains moral disagreements as attempts by parties at spotting inconsistencies in each other's claims. That possibility is blocked by our main counterexample to constructivism, the

Ideally Coherent Caligula (ICC). We must resist him on other grounds, for he is perfectly consistent. But disagreements assume other forms. Sometimes they manifest themselves as metalinguistic negotiations. In that spirit, we came up with a possible conversation with the ICC, where a representative of some standard Western morality metalinguistically negotiates with him. That is likely to be part of what we do when we resist accepting the ICC is not making any mistake. We further saw that there is some empirical evidence pointing in the direction of that acceptance, particularly when parties are culturally far away from each other. Thus, it is reasonable to expect that participants in a hypothetical experiment would consider both the average Westerner and the ICC correct in their judgments, though obviously the experiment would have to be run to assess this hypothesis.

The empirical evidence aforementioned doesn't come from analyses from the point of view of the disagreeing parties, but rather from the point of view of an external assessor, to whom a report is given about two disagreeing individuals. Therefore, we are not entitled to extrapolate from the results of these studies the idea that disagreeing parties themselves would be more inclined to perceive their opponents as being correct the more culturally removed from each other they are. However, the final alternatives I wish to highlight from now on predict precisely this kind of extrapolation.[70]

The point I would like to stress is the following. If a party (A) to a value disagreement is not a) checking her opponent's (B's) consistency; b) metalinguistically negotiating with B (or adopting other similar strategies); or c) simply passively accepting the disagreement she takes part in (as the evidence suggests that often is the case); then, assuming that A's action is justified, I argue that most likely *she is engaged in attempting to modify B's behaviour and/or values*. Of course, A could be acting irrationally, randomly messing with B's behaviour. But we are hypothetically ruling out this possibility. What I wish to explain is the *justified* resistance by A in reaction to B's values or behaviour, even after acknowledging that B is not making any mistake from B's point of view.[71]

[70] Though the data don't provide direct support for this extrapolation, they certainly don't suggest anything against it. It is an empirical matter that needs to be verified.

[71] At this point it is important to bear in mind that constructivism conceives of moral and other evaluative judgments in an internalist way, that is, from the point of view of the agent uttering those judgments, as explained in chapter 1 and at the beginning of this chapter.

The general claim I want to make is that A's justified reactions are a function of the normative relevance of B's actions to A. For this reason, A persists in criticizing B's values and behaviour and persists in attempting to modify them to the extent that B's values and behaviour generate consequences that conflict with A's values.

As stressed at different points in the previous chapters, for constructivism the justification for actions and moral judgments asserted by an agent is assessed by taking that same agent's set of values as the standard. Thus, B's values are the source of justification for B's action, and, for that reason, A is interested and justified in attempting to modify B's values. But it is primarily B's actions (or the manifestation of B's values in his actions) that matter to A. Therefore, A is primarily engaged in modifying B's behaviour and not so much his values. Let's explore some of these kinds of justified reactions that A might manifest in relation to B, in the face of mutually acknowledge faultless disagreement.[72]

Sometimes A is in the business of blaming B for what A takes to be something wrong that B did. However, blame typically requires that the agent failed somehow to fulfil her obligation and, in a case of faultless disagreement, we are assuming that B acted in accordance with his reasons. Bernard Williams has suggested that there is a kind of blaming attitude an agent might engage in that is not dependent on the blamed party having failed to do what he had reason to do. This kind of blame rather targets a more general reason that is part of the blamed one's set of values[73], namely, his willingness to be seen as a person worthy of respect by the blaming party. Appealing to this more general reason, the blamer wishes to *produce* a reason in the blamed party, a reason that was not there before, but that, if the blaming is successful, will emerge out of the blamée's recognition that he failed to meet the expectation of the blamer, whose esteem he values (Williams, 1995).[74]

Williams' story works for some cases, but not for others. The ICC and other similar subjects don't count among their values a willingness to be esteemed by others (*ex hypothesi*). It has been suggested that yet another kind of blame is appropriate for those who haven't done anything subjectively wrong. Mason (2019) describes what she has called detached blame, a blaming attitude that is not capable of speaking to

[72] The notion of faultless disagreement is due to Max Kölbel (2004).
[73] Williams would have said: part of the agent's subjective motivation set. See Williams (1995, 1979).
[74] Miranda Fricker has acutely explored this kind of blame and its importance for the construction of a shared moral community in her 2010 and especially in her 2016 (Fricker 2010, 2016).

the blamed party's reasons or values.[75] As Mason rightly points out, often the point of this kind of non-communicative blame (whose purpose is not to offer communicatively reasons to the blamée) is not to affect the blamed party, but to signal to other members in our community our commitment to our values and their upholding (2019: 122). Maybe blame is not the appropriate word for the attitude, but the point is that it is a completely understandable reaction, with concrete consequences (if not for the blamed party, at least for others relevant to the blamer).

It is worth referring to a classical distinction drawn by P. F. Strawson (1974) to better understand what we have discussed up until now. Strawson distinguishes between two attitudes we might adopt in relation to other people: a participatory attitude and an objective attitude. In adopting the participatory attitude, the most common one, we get involved in interpersonal relationships characterized by, say, affection, diverse reactive attitudes, exchange of reasons and arguments, etc. In contrast, by adopting the objective attitude, these modes of relationship are broken and we take up a detached posture regarding other people. We observe them as objects of the world, not as people to whom reasons can be offered or with whom we might establish some kind of affective or emotional contact. As Strawson put it, "if your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him" (1974: 10).

Back to our previous argumentative thread, we can surmise, in the first place, that in moral disagreements parties generally adopt a participatory attitude towards each other, whereby they are in principle willing to engage with each other's reasons. This is generally what happens when parties attempt to spot inconsistencies in the arguments and views of their interlocutors. In a second moment, when disagreement reveals itself as faultless but persistent, parties might adopt what we can call an intermediate stance between the participatory and the objective attitudes. In other words, parties may seek to blame their interlocutors along the lines proposed by Bernard Williams. In that case, the blaming party is not engaging with the particular reasons that make her interlocutor utter her judgment and, because of that, doesn't

[75] Mason's model works under the assumption of an objective standard – Morality, as she calls it – with which to assess whether an agent's actions are objectively wrong (in contrast with subjectively wrong, that is, wrong from his point of view). If objectively wrong, actions are, for Mason, adequate objects for blame. However, the point about detached blame is also applicable in the context of a relativized conception of practical normativity.

treat her as a person in the "space of reasons" anymore. On the other hand, the blamer looks for other reasons the blamée has to try and change her mind. Moreover, this kind of blaming is only effective so long as the blamed party possesses a desire to be recognized as worthy of the blamer's esteem. Both these last two elements are characteristic of the participatory attitude as described by Strawson. Such cases are, thus, intermediate.

In the case of detached blaming, discussed by Mason, the objective treatment of the interlocutor gains prominence, in that the blamer realises that engaging with her interlocutor's reasons is no longer an effective strategy to bring about her desired outcome. The objective attitude in face of persistent and radical moral disagreement is a limiting attitude, but it is at times justified. When? Reflecting on this question, I bring to an end my considerations about the different justified reactions to moral disagreement, according to constructivism.

We saw that blaming is one of the things agents can do to others in contexts of faultless disagreement. Another kind of reaction an agent (A) might be inclined to have when faced with someone (B) with deeply contrasting values in a situation of faultless disagreement is simply try to stop B from doing something that A finds, according to her values, to be wrong. One might ask what *warrants* A's behaviour? With what *right* does she try to prevent B from doing what he *has* most reason to do according to his values?

Well, part of the answer has already been given for the case of detached blame. A is signalling and in fact manifesting her commitment to her values (the values of her community). If that is true for blaming B, it can also be true for behaviour that tries to stop B.[76] But I believe this point can be given a more general explanation and support.

In a situation of faultless disagreement, over and above the conflicting doxastic attitudes in relation to the same proposition, sometimes one of the parties (A) will engage in a series of actions aimed at modifying the other's (B's) behaviour in ways that make it compatible with the realisation of her values (A's).[77] This series of actions will quite often be justified, for *they are part of the agent's implementation of her values*, that is, of her practical reasons. In other words, what warrants A's attempt to modify B's behaviour (in a situation in which otherwise B would act in accordance

[76] Maybe just some mild counter movement is justified along these lines.
[77] Both parties can do this simultaneously.

with B's reasons) is that A's attempt at modification of B's behaviour is an action itself. It is an action by A that ought to be assessed by reference to A's standard (namely, her set of values), just like any other action. This is constructivism's internalism, to which I alluded earlier. Thus, in general whenever B's behaviour produces or is about to produce in the world a state of affairs that is contrary to what A values, A has a reason to prevent B from doing what he is nevertheless justified in doing.[78]

If this approach is correct, it exactly predicts that the farther away B's behaviour is from A's sphere of influence or space of action, the less A will be bothered by it. For it will, *ceteris paribus*, interfere less with the implementation of her values. With this explanation, that is, resorting to judgment internalism as is characteristic of constructivism, I believe I contribute to the understanding of why doxastic disagreement seems, to participants in the experiments surveyed, to be less problematic when the parties are greatly separated.

Thus, to return to our guiding example in this chapter, any typical Westerner that encounters an ICC will most likely have every reason to act, in the most diverse ways, so as to stop him from putting his life policy into practice. This is so because, in all likelihood, the implementation of the ICC's life plan will amount to a major obstacle to the implementation of that Westerner's values. And the latter, as seen from her point of view, will have every reason to prevent that from happening. Under such extreme circumstances, disagreement takes up the form of resistance and the objective attitude imposes itself.

***

Accounting for disagreement with extreme characters like the Ideally Coherent Caligula requires taking a complex view on the matter. The analysis must accommodate the different kinds of strategies the ICC's opponent might adopt while disagreeing with him. This chapter covered four of them: checking inconsistencies; engaging in metalinguistic negotiation; passively accepting the genuine, doxastic disagreement (as suggested by evidence from empirical studies); and engaging in attempts, via blame or action, to shape the ICC's values or practices.

[78] If this is correct, it is a possible response to Boghossian's worry: "Further, I'm trying to get you to change your mind – why would I do that if I regarded your standards to be just as correct as mine." (2011: 64).

Thus, the chapter has attempted to make two kinds of contribution so far. First, by coming up with this list of possible reactions to the ICC, it strengthens the case for constructivism and reduces the room for resistance to it. Second, by introducing the fourth strategy – blaming and counter-acting – it explains in a natural way what the other strategies might fail to capture.

In the remainder of the chapter, I will introduce another kind of problem for relativist theories of morality. In order to properly tackle the critique, I will have to resort to some of the recent literature in the philosophy of language and disagreement. In the first place, my ambition with the following remarks is to offer further support for constructivism about practical reasons. Secondarily, however, I intend my arguments to lend support to a specific theory in that debate in the philosophy of language, namely non-indexical contextualism. This is because constructivism, if I am correct, helps explaining part of the moral practice that is not dealt with in the philosophy of language. Therefore, I shall claim that constructivism and non-indexical contextualism help each other in perhaps unexpected ways.

4.7. Philosophy of Language and Moral Disagreement

Some philosophers believe there is no such thing as (intercultural) moral disagreement.[79] However, although highly stylized, there seems to be nothing impossible about a disagreement between an average Westerner and the ICC. If that is the case, accounting for such disagreement is one of the things we are entitled to expect from a moral theory.[80] Furthermore, the way I see it, the case of the ICC is mostly a useful theoretical device that helps us grasp some of the elements of what is at stake in moral disagreements in pure form, that is, deprived of the inconsistencies that plague all real persons. Earlier in the chapter we saw that there might be a thin empirical support for the idea of folk moral relativism. In this last section, I want to examine the suspicion that moral relativism is philosophically impossible to sustain, that is, that it is an incoherent position.

So how do we explain moral disagreements with a little more detail? Most authors traditionally called relativists in metaethics can be classified as what is now

[79] E.g. Velleman (2013: 25).
[80] Street (2009) also acknowledges the importance of accounting for this kind of case.

called, under the influence of philosophy of language, indexical contextualists (IC).[81] In its simplest form, moral indexical contextualism is just like contextualism for indexical terms like "today". So, if I say, on 5th October 2015:

3) *Today is sunny.*

And you say, on 5th October 2016:

4) *Today is not sunny*.

We are not disagreeing so much as we are talking about different things. If we substitute "today" in 3 for "5th October 2015" and "today" in 4 for "5th October 2016" it becomes immediately clear that 3 and 4 are not incompatible assertions. According to IC about morality, the same happens with moral terms. So when Immanuel says:

5) *Lying is always wrong.*

And Jeremy says

6) *Lying is not always wrong*

The terms "wrong" in 5 and "wrong" in 6 mean different things. On one interpretation, they are abbreviations for "wrong according to Immanuel's standard" in 5 and "wrong according to Jeremy's standard" in 6. It becomes clear, then, that Immanuel and Jeremy are not engaged in asserting two mutually excluding sentences. Rather, their assertions are perfectly compatible and both might very well be correct.

This theory is perfectly suitable for so-called relativists who claim that moral disagreement is impossible, as mentioned before. Other defenders of IC will strive to account for disagreements in terms of metalinguistic negotiations or other kinds of non-doxastic disagreement.[82] Although there are multiple strategies available for the IC, it seems to me that, at least sometimes, disagreements are doxastic, i.e., about conflicting contents. Referring specifically to the contents asserted by the disagreeing parties, then, let me present the problems with indexical contextualism with a little bit more detail.

First, there is a problem we might call the problem of the loss of normativity. When X says "Child abuse is wrong" and Y says "Child abuse is not wrong" they are both uttering sentences with normative content. When the indexical contextualist translates X's sentence to "Child abuse is wrong according to X's standards" the translated sentence is not a normative sentence anymore, but rather merely a descriptive sentence about the components of X's standard or a merely logical claim

[81] For an introduction to the discussion that follows, see Stojanovic (2017).
[82] See (Zeman 2017) for an overview of contextualist strategies.

about what follows from X's moral parameter. As Boghossian (2011: 58) clearly expressed, on this view we seem unable to explain how there could be such a thing as a disagreement about normative matters (Boghossian 2011).

Another problem for the IC, also diagnosed by Boghossian, is that on this picture,

> there is nothing very exciting going on as far as truth is concerned. Contents have absolute truth-values. The only sense in which anything is relative to anything else is that the thinkable content expressed by a token of a sentence type is relative to that token's context of utterance (2011: 59).

In other words, on this view, the content of all moral claims is relative, but their truth-values are always absolute. Assessors should refer to the context of utterance of a given token of a sentence to assess its truth-value and all assessors should agree on the verdict. This is another way of saying that disagreement about content is impossible on indexical contextualism. Being so, this justifies looking for alternative views.

A closely related view allows for doxastic disagreement. According to Non-Indexical Contextualism (NIC)[83], when Immanuel asserts 5 and Jeremy assert 6 they are genuinely disagreeing. On this view, Immanuel and Jeremy hold opposing views about the same proposition. So, while in IC, the content of the proposition is contextually defined in reference to the speaker's standard, in NIC the content of the proposition is fixed. So if Immanuel asserts 5 and Jeremy asserts the opposite of 5 (=6), they are talking about the same content. In the words of Berit Brogaard, "as the contents of 'right' and 'wrong' are context invariant, there is something for the disputants to disagree about, *viz* the invariable content the truth-value of which depends on the standards of the judge" (2008, 393).[84]

NIC allows for doxastic disagreement and, at the same time, claims that both parties in the disagreement might be correct. When this happens, they are engaged in what Kölbel (2004) called a faultless disagreement.[85] The disagreement can be

[83] I adopt the terminology found in MacFarlane (2014, 2009).

[84] The quote from Brogaard characterizes both non-indexical contextualism and assessor-relativism, the view we will discuss next, because it doesn't specify if the judge in question has to be the agent or speaker (as in non-indexical contextualism) or if she could be a third-party assessor (as in assessor-relativism).

[85] "A faultless disagreement is a situation where there is a thinker A, a thinker B, and a proposition (content of judgement) p, such that: (a) A believes (judges) that p and B believes (judges) that not-p (b) Neither A nor B has made a mistake (is at fault)" (Kölbel, 2004: 53-4). See also Berit Brogaard's definition of faultless disagreement in contrast with objective disagreement: "I shall say that a

faultless because the standard of evaluation for each party's assertion is different. So it might very well be the case that, according to Immanuel's standard, 5 is justified, while according to Jeremy's standard 5 is not justified. When that is the case, an assertion of 5 by Immanuel and a denial of 5 by Jeremy constitute a faultless disagreement. According to this view, each agent's perspective does not step in at the moment of determining the content of what is asserted, as in IC. Rather, the agent's perspective is the function assigning truth-values to propositions (Kölbel, 2004: 70). Assertions of the same proposition by different agents may have different truth-values.[86]

Several authors have resisted the idea that disagreements such as the one just described could be coherently called faultless (see Boghossian, 2011; Richard, 2008; Wright, 2008). They all more or less point to a disquotational schema. Their argument is roughly the following. Assume that non-indexical contextualism is right and that the truth of normative assertions is only relative truth (truth relative to the perspective of the utterer). Even then, these critics claim, it must be valid of someone that judges p that she can say: "It is true that p". But if she can say "It is true that p", she can also say "It is false that not-p". But if she can say "It is false that not-p" she cannot coherently claim that anyone judging that not-p is not making a mistake. Hence the impossibility of faultless disagreement.

Commenting on an alleged faultless disagreement between D and N and referring to a group of views (Alethic Relativism) of which NIC is one specimen, once more Boghossian lays out the argument with precision:

> (13) The content (p) is at best relatively true. (Alethic Relativism)
> (14) If D judges validly that p, it will also be valid for D to judge that It's true that p. (Truth is Disquotational within a perspective)
> (15) If D judges that It's true that p then D must, on pain of incoherence, judge that It's false that not-p.

---

disagreement is objective iff (i) there is a proposition p whose truth-value is the subject of disagreement; (ii) relative to each of the disputants' circumstance of evaluation i, one of the disputants assigns the incorrect truth-value to p and the other assigns the correct truth-value to p; and (iii) relative to each of the disputants' circumstance of evaluation i, each disputant x would assign the truth-value x actually assigns to p, had x been in a context which determined i. A disagreement is faultless iff (i) and (ii) are satisfied but (iii) is not." (2008: 392-3)

[86] In unpublished work, Street assumes indexical contextualism and is, thus, led to the adoption of the problematic view that not only relativism is a good theory of normative reasons but that we should also be relativists about this claim, i.e., that we shouldn't be absolutists about the truth of relativism about normativity. This creates several complications for her view, which could be largely avoided if she adopted non-indexical contextualism.

> (16) If D judges that It's false that not-p, then D must, on pain of incoherence, judge that anyone who judges not-p (e.g., N) is making a mistake.
> Therefore,
> (17) D must judge that N is making a mistake and so cannot regard the disagreement with N as faultless.
> Therefore,
> (18) The disagreement between D and N is not faultless. (2011: 62)

The argument is powerful but I believe it can be adequately countered. I will hint at two alternative responses, without the ambition of having exhausted the matter. The first alternative, due to Berit Brogaard (2008), contradicts directly the argument. The point is to recall that this argument is presupposing that the default judge of a given truth-predicate is always the evaluator and not the speaker, as non-indexical contextualism would suggest. For, remember, what the example above shows is that D cannot simultaneously judge that not-p is false *and judge* that N is not making a mistake in asserting not-p. Brogaard's argument is to recall that, according to perspectivalism (her version of non-indexical contextualism),

> the truth-predicate means something different when it is restricted to the meta-linguistic level from what it means when it is not so restricted. When it is not so restricted, we can infer 'John is a firefighter' from '"John is a firefighter" is true'. But we cannot do this when it is restricted. When it is restricted, the utterance 'John is a firefighter' may be true if John is a firefighter from 1990 to 2000. So 'The utterance "John is a firefighter" is true' does not entail that John is a firefighter. In other words, some difference must exist between meta-linguistic uses of sentences where the truth-predicate is restricted to the meta-linguistic level and uses where it is not so restricted. (2008: 407)

And she next claims, that

> the truth-predicate that is restricted to the meta-linguistic level is the usual one (given Kaplan-style semantics); *s* as uttered in *c* is true *simpliciter* iff the proposition expressed by *s* in *c* is true at the circumstance of evaluation determined by the context of the speaker who uttered *s* (and not the context of the semanticist). In other words, the main difference between [assessor] relativism and perspectivalism concerns meta-linguistic uses of sentences where the truth-predicate is restricted to the meta-linguistic level. (*Ibid*: 407-8)

The details of Brogaard's view are complex and need not detain us here. Her general point is that the validity of the disquotational schema that translates from a judgment that p to a judgment that p is true and, thus, grounds the whole argument of the critics of non-indexical contextualism is based on a presupposition

that the truth-predicate in the analysed cases is not restricted to the meta-linguistic level. And, contrary to that, she claims that restricting the truth-predicate to the meta-linguistic level is the usual attitude within the framework of standard Kaplanian semantics.

Another approach to the criticism above is to accept it, but reduce significantly its import. The idea is to acknowledge that, for disquotational reasons, it is inappropriate, in the example above, for D to say that N's assertion of not-p is true. However, the suggestion now is to say that refraining from granting your interlocutor's assertion the status of "true" (or, *mutatis mutandis*, "false") doesn't amount to claiming that she is making a mistake. So, perhaps the notions of truth and falsehood cannot be used the way non-indexical contextualists originally thought they could, but surely some other notion, like accuracy, can do the work. In this proposal, D is unable to say that N's assertion that not-p is true, because D's use of the word "true" picks out necessarily D's standard, but D *is* able to say that N's assertion that not-p is accurate, if we postulate that accuracy is meant to pick out what is true according to the standard of the speaker, not the evaluator. Thus, if both D and N are able to see each other as accurate when asserting respectively p and not-p, then it seems like it is after all appropriate to speak of faultless disagreement, just like non-indexical contextualists recommend.[87]

Now, non-indexical contextualism is the view I endorse. However, it is worth briefly mentioning another alternative theory and sketch an argument for not embarking on it. In a series of papers (MacFarlane 2005, 2007, 2009), culminating in his 2014 book, John MacFarlane has proposed what he called Assessor Relativism (AR). Remember that IC arguably has problems with accounting for disagreements. For MacFarlane, NIC does not fare much better in this respect and this is so for reasons different from the ones we discussed so far.

Unlike IC, NIC accounts for the kind of disagreement that exemplifies what MacFarlane calls the condition of Noncotenability, the idea that I disagree with

[87] This final proposal might help the non-indexical contextualist in yet another way. Allegedly a problem for non-indexical contextualism comes from data from experimental philosophy roughly showing that people use words like "false" and "no" when opposing statements made by interlocutors whose standards of evaluation predict the opposite of what their standards (of the ones who react saying "false" and "no") would predict. By allowing faultless disagreement to coexist with coherent and justified attributions of falsehood to the assertions of the disagreeing party, non-indexical contextualism comes up as not obviously contrary to the available empirical evidence. See Khoo & Knobe (2016) for the empirical study and a defense of indexical contextualism based on data such as the above.

someone's doxastic attitude if, in order to incorporate her attitude, I would have to change my mind (2014: 121). However, MacFarlane claims there is a further sense of disagreement that NIC and IC are both not capable of accommodating. That is a disagreement fulfilling the condition of Preclusion of Joint Accuracy: "The accuracy of my attitudes (as assessed from any context) precludes the accuracy of your attitude or speech act (as assessed from that same context)" (2014: 129). In NIC, joint accuracy is not precluded. It is possible for A to be accurate in claiming that p, while B is accurate in claiming that not-p. Regardless of whether A, B or a third party assess A's and B's claims, both claims may be deemed accurate from the point of view of the same assessor.

Contrary to this, AR endorses the Preclusion of Joint Accuracy. For AR, the accuracy of a claim is always evaluated from the point of view of the assessor. So, to repeat the example above, if A claims that p and B claims that not-p, it is impossible for both to be accurate. For, if A is the one assessing the accuracy of both claims, A will do so with reference to her perspective. If, on the other hand, it is B who is assessing both claims, then B will do so with reference to his perspective. The perspective of reference is always the perspective of the assessor and, therefore, it is impossible for joint accuracy of incompatible claims to obtain.

Fully responding to this challenge would require detailed argumentation. I only sketch three arguments for sticking to NIC in the face of MacFarlane's critique. First, I am not fully convinced of the relevance of Preclusion of Joint Accuracy. It seems to me that Noncotenability captures the essence of what is going on in moral disagreements (apart from the other aspects discussed in previous sections). Noncotenability captures the idea that parties in a disagreement express or hold opposing attitudes towards the same content and it is not obvious why we want anything more than that.[88]

Secondly, assessor-relativism adds an extra layer of semantic complexity. As Brogaard notes, assessor relativism "relativizes sentence truth to a context of use and a context of assessment. As a result, what is said on a particular occasion does not have an absolute truth-value at the context of use." In contrast to that, her view (which she calls perspectivalism, but is really a version of non-indexical contextualism)

[88] I conjecture that wanting more of disagreements, that is, demanding that the opposing party cannot also be accurate moves beyond simply being in a state of disagreement and goes into the territory of what parties might be willing to do when engaged in a disagreement. This is the topic of the previous section of the chapter.

"does not relativize sentence truth to contexts of assessment; it relativizes only to contexts of use. So what is said on a particular occasion has an absolute truth-value with respect to the context of that occasion." (2008: 404). According to non-indexical contextualism, the truth-value of a particular normative judgment doesn't change every time a new assessor assesses it. In assessor-relativism it does.

Thirdly, and finally, it is worth remembering the internalist feature of constructivism I mentioned above. We are looking for the best relativistic account of moral disagreement that matches the constructivist framework and, as we saw, constructivism explicitly conceives of normative justification *from the standpoint of the agent*. That is, an external assessor, according to constructivism, is expected to evaluate a given agent's values and actions from *that agent's* perspective. I consider the fact that an elaborate view about practical reasons such as constructivism is readily available to NIC as a point in its favour.

In sum, if constructivism is the correct theory about practical reasons this is good news for non-indexical contextualism. And this is so not only because of the internalist features both views share. For example, the blaming and counter-acting attitudes I discussed above make it possible to rationally justify, with appeal to constructivism, certain critical reactions to the ICC even in the face of mutually acknowledged faultless disagreement. In other words, non-indexical contextualists wishing to account for the complexities of real-world moral disagreements will not have problems explaining why such resistant reactions occur in cases where their theory predicts faultless disagreements. For they now have a normative theory at their disposal, namely constructivism, that explains and occasionally justifies such reactions.

5. Setting Limits to Practical Reflection (Against Philosophy as a Way of Life)

In the previous chapter I examined some of the potentially unwanted consequences of the version of constructivism about practical reasons that I favour. We saw that, by accepting relativism about value, the view is open to counterexamples such as that of the Ideally Coherent Caligula (ICC). Hopefully I was able to disperse some of the reader's concern with the arguments I presented then to the effect that figures such as the ICC don't pose that much of a threat to relativist theories of practical normativity such as my preferred version of constructivism.

In this chapter, I proceed with the exploration of the consequences, both practical and theoretical, of the view I have been endorsing along this dissertation. This time the attention is directed to a reasonable expectation someone might have after being acquainted with a broadly constructivism view. In fact, constructivism, in its different guises, apparently gives pride of place to reflection or deliberation in the establishment of what a given agent has reason to do. In that, the view is loosely reminiscent of an old tradition that considers (self-) examination – or even the philosophical life, in particular – a precondition for the life properly lived. Next, I will explore this loose parallel and, more importantly, evaluate the pros and cons of that life policy and to what extent constructivists should endorse the programme of philosophy as a way of life.

5.1. Philosophy as a Way of Life: Socratic Style

Human beings are typically agents. We act for reasons, frequently out of a conscious apprehension of such reasons. As rational beings, we also have a capacity to reflect on the considerations we take to be reasons, to scrutinize them.[89] In very crude form, the philosophical way of life, as proposed by Plato's Socrates, means submitting all our alleged reasons and beliefs to the test of philosophical examination.

[89] Following Scanlon (1998) it has become commonplace to explain reasons as considerations that count in favour (of some action or belief). As Velleman (2009: 121-2, fn. 8) correctly notes, simply saying that a consideration counts in favour of something doesn't explain much. We want to know *why* it counts in favour of whatever it counts in favour of. In other words, we want to know how it *warrants* or *justifies* that action or belief. Therefore, a reason is best understood as a consideration that justifies (*Idem*: 122). The reflective scrutiny referred to in the main text is a pursuit for such justification.

The bet is that a life conducted along these lines has a better shot at making us wiser and more capable of achieving the good.[90]

Socrates famously claimed that "the unexamined life is not worth living" (*Apology* 38a 5-6).[91] The passage is most likely best understood not as claiming that an unexamined life is worse than being dead, but that one *should not* live an unexamined life.[92] Socrates' self-examination and examination of his interlocutors (so-called cross-examination) is deeply connected to his ethical enquiry, to his efforts to lead a good, virtuous life. For one thing, self-examination can help extirpate the mistaken conceptions and assumptions one inherits from culture and unreflective education.[93] Under the supposition of an objective conception of the good life, thorough, philosophical self-scrutiny can also help put one on the right track, on the track of wisdom.

Socratic wisdom is attained when one has fully grasped the truth about all human values and their systematic relations. Virtue is the condition of the soul of someone possessing wisdom. The wise and virtuous person always does what is best, for her actions always spring from her knowledge of the good (Cooper, 2012). Therefore, constantly seeking wisdom is apparently fully justified within the Socratic model. For, it disposes us of false preconceptions and puts us on the way to a happy life.

Socrates was convinced that achieving this kind of wisdom was beyond the capacity of any real human being. Although wisdom was for him the best life in principle, in practice the best life was philosophy, the love of wisdom, or its constant pursuit. Committing to philosophy as a way of life was committing to following reason wherever it may lead .[94] As Cooper puts it, pursuing wisdom requires "constant philosophical discussion about matters of human value, and […] constant self-examination of one's own views on the fullest range possible of those questions" (2012: 57).

---

[90] Discussion about philosophy as a way of life among the Ancient was pioneered by Pierre Hadot (Hadot 1995). For the Ancient, according to him, a life dedicated to the pursuit of wisdom meant a life that brought peace of mind, inner freedom and (for the Stoics and Epicureans) cosmic consciousness (Hadot 1995: 265-6). Hadot's emphasis on spiritual progress has been criticized by John Cooper (2012). I do not wish to engage with this polemic. I register, though, that my presentation in the text follows more closely Cooper's reading.

[91] I quote Plato according to Cooper's translation (Plato 1997).

[92] Kraut (2007) proposes to translate the passage as: "no human being should live an unexamined life" (2007: 231). I follow his interpretation in the next phrases in the text.

[93] Kraut makes this point referring to many of the so-called Socratic dialogues (2007: 238).

[94] Cf. *Apology* 29 c-d for Socrates' persistent commitment to philosophical enquiry.

An immediate thing to notice about this model is how demanding it is. It has at least three noteworthy features. (1) It requires that the agent fully examine her values and beliefs about several things; presumably all her views about what she has reason to do. (2) It requires that she know all there is to know about matters of value and reasons for action, at least when it comes to moral decision-making. (3) And it presupposes that these kinds of (self-) examination and knowledge are the best guides to the happy life. Because achieving (2) is admittedly almost impossible, the follower of the model risks becoming obsessed with (1), in the expectation of reaping the benefits promised by (3). No wonder why the Socratic model was met with a great deal of suspicion and scepticism since the beginning.

Perhaps Socrates' model is to be taken precisely as a model, that is, as an ideal, which we shouldn't expect real human beings to fully embody. The ideal model could, then, be seen either as a provider of ultimate practical reasons for actual human beings – for the model would not be plagued by the imprecisions of actual cases – or as a point of reference, to be approximated as much as our human nature allow.[95]

However, independently of how difficult it is to adopt the Socratic way of life, what I wish to investigate in this chapter is whether we have reasons to strive for such a life in the first place. In other words, I investigate whether we are justified in questioning our reasons in a self-examining process looking for firm foundations. I respond to these questions in the negative and the better part of the chapter will be dedicated to arguing for the claim that there are more or less precise limits to how much practical reflection[96] is warranted in each situation of action. Just like talking or walking, deliberating is an action and, as such, requests a warrant before one engages justifiably in it. Very often one does not have reason to (further) deliberate – or so I will argue.

[95] Idealized models have diverse problems of their own. Just to mention one: it seems that agents would have to have reasons to comply with the choices of their idealized counterparts.

[96] Practical reflection, otherwise called practical reasoning or deliberation, is reflection about what to do. Socratic self- and cross-examination is of course concerned with the knowledge of truth in general and for that reason doesn't limit itself to reflection about reasons for action. In this regard, my discussion in this paper will have a more limited range. Notice, however, that interrogating our beliefs is arguably also something we do and whose justification can just as well be cast into doubt by the exercise of our practical reasoning capacities. If that is true, my discussion here will have implications with similar breadth as that of Socrates. But this is a matter I wish to remain neutral about.

5.2. Anti-realist constructivism

In this section, I briefly recall the metaethical view with which I will be working in this chapter. I am not going to argue for its comparative merits here.[97] It is nevertheless important to have a clear starting point.

The view is called metaethical constructivism. I believe the best general definition of it is given by Sharon Street:

> According to **thoroughgoing** or **metaethical constructivist views**, the truth of a normative claim consists in that claim's being entailed from within the practical point of view, where the practical point of view is given a *formal* characterization. (2010: 369)

When an agent makes a normative judgment, say claims that X is a reason for her to do something, constructivism offers a model to evaluate whether the agent is correct in her claim. The agent occupies what Street calls the practical point of view when she makes normative judgments, takes some consideration to be a reason or, more generally, values some things over others. Her practical point of view is in a sense composed by her set of values[98], based on which she makes her particular normative judgments. Makes or should make. In fact, a normative judgment will be justified as long as it is entailed from that practical point of view, from her set of values. The practical point of view is given a formal characterization because it is not meant to have any substantive assumptions. It is the standpoint of the agent as such, regardless of which particular values a particular agent holds.[99] The notion of entailment is not defined with complete precision, but is meant to include instrumental rationality, logical entailment and conceptual necessity.

The view is internalist in at least two senses. It is a kind of motivation-internalism, in that the notion of valuing or taking something to be a reason is expected to (at least partially) motivate the agent. So, for example, if an agent fails to be motivated by what she acknowledges to be a necessary means to her valued end, she either doesn't really count as taking it as a reason or is instrumentally irrational (Street 2008: 230). It is also a kind of judgment-internalism, where it is to be

[97] I present the view with more detail in chapter 1 and defended it from critics at different point of the dissertation, most notably in chapters 1 and 4.
[98] The identification of the agent with her total set of coherent values is the view I defended in Chapter 3.
[99] In Chapter 1, I take issue with the details of Street's formal characterization of constructivism, but that doesn't have to concern us here.

determined whether a consideration is a proper reason by seeing things from the point of view of the agent. What is considered externally to be a reason for that agent doesn't qualify as a reason for her if it doesn't gain that status of a reason from within her point of view.

Moreover, on the constructivist view, justification is coherentist. A particular normative judgment will be justified if it coheres with the total set of values possessed by the judgment's utterer. No individual has a perfectly coherent set of values, so it becomes a matter of gradation how coherent the set is. Also, a particular judgment can have more or less connections to the other judgments, values and beliefs in the set. More connections mean more justification. Finally, particularly for the case of values and normative judgments, the more deeply attached the agent is to the value or judgment – which may or may not supervene on the level of connectedness of the value or judgment with the rest of the set – the more the judgment or value is justified. Deeply held values are closer to the core of the agent, they are more defining of her identity and, therefore, she has less *prima facie* reason to abandon them.

What is the agent's total set of values and beliefs? Which ones cohere with one another? Which dispositions to believe and value does the agent possess that might be inconsistent with the normative judgment she just made? More generally, what is entailed from her practical point of view – her total set of values – in conjunction with the non-normative facts of a given situation so that she can know what she ought to do? Answering all of these questions seems to require that the agent engage in self-examination and practical reflection, casting doubt on what might have appeared to her as a reason for action. The reflective, philosophical way of life slowly suggests itself.[100]

There are important parallels and contrasts to be drawn between the constructivist model outlined in this section and the Socratic model discussed above. The major contrast derives from the fact that Socrates presupposed a realist conception of the good and construed philosophy as the reasoned attempt to direct

[100] In fact, Korsgaard, a leading, Kantian constructivist, has expressed herself clearly about the high value she places on reflection in the following terms: "Kantian positions in general set a high value on reflection and are idealizing positions in the sense that moral concepts, as Kant defines them, are derived from the ideal of a fully reflective person. The fully reflective person is a corollary of Kant's idea of the unconditioned. We seek the unconditioned by imagining a person who reasons all the way back, who never gives up until there is a completely undeniable, satisfying, unconditional answer to the question. Obviously human beings often stop reflecting very far short of that. And reflection itself is not the solution to that problem. So in that sense reflection is not the complete guide although it is the only place where we can find guidance. Something else has to get us to reflect." (2003, 60)

oneself to the independent good. Contrary to that, constructivism sees all value originating from human agency and the practical point of view. In terms of similarities, although contemporary constructivism is more explicit about it, both models highlight the importance of coherence to belief and value justification. Socrates recommends knowing the whole system of values and their systematic relations and interconnections.[101] Constructivism explicitly endorses coherentism. Finally, and more importantly, both models seem to require a kind of reflective scrutiny on the part of the agent, before she is justified in taking some consideration as a reason for action. In other words, both models seem to posit regular practical reasoning as the only way to living in accordance with one's reasons. This last feature is in all likelihood true about Socratic philosophy. What I will attempt to demonstrate in the remainder of this chapter is that it is not necessarily true of metaethical constructivism. This is a good conclusion for constructivism, for, as I will also argue, the fully examined life is not a justified life policy.

5.3. Limiting Practical Reflection

There is a convincing case to be made against considering reflection or deliberation necessary for autonomous rational action. In fact, in many cases we seem to have no reason to engage in deliberation, cases in which we do what is rational to do "in the flow", out of habit, or in a kind of "fluent agency".[102] I do not wish to discuss the literature on these cases here. If proven correct, it provides arguments to the effect that often we should not deliberate *at all*. That is a good result for me. But I wish to grant to the defender of Socratic full-examination that we are considering only cases of conscious, deliberate, rational decision-making. Limiting the discussion to cases of conscious deliberation, what I wish to confront the Socratic philosopher with is the idea that there are limits to what we should consider and to what we should challenge. In general, I claim we shouldn't adopt a kind of Socratic stance towards our values and beliefs.

[101] Suspending my judgment on its plausibility from the scholarly point of view, it is worth mentioning Vlastos' classic interpretation of the Socratic method of refutation. Vlastos ascribes the following principled belief to Socrates: "Whoever has a false moral belief will always have at the same time true beliefs entailing the negation of that false belief " (Vlastos 1994, 25) . From this assumption, it is possible to extract the thought that a person with fully consistent beliefs will, by necessity, have only true beliefs. Hence the importance of verifying one's consistency, according to this interpretation.
[102] Cf. (Kornblith 2010; Railton 2009; Arpaly and Schroeder 2012)

In the case of reflective decision-making, the agent is consciously taking some consideration (or set of considerations) to be a reason to do something. What warrants or justifies her taking that consideration to be a reason? Perhaps there are substantive facts, natural or non-natural, about what counts as reason for action. But as a metaethical constructivist, I want to assume away normative realism. The pressing case for the need to come up with a non question-begging explanation for *why* a certain consideration counts as a reason for action has been convincingly put forth by Velleman (2000b).

Velleman's considered view is that the ultimate criterion of justification for an action is whether the action makes sense for the agent to perform.[103] Now, Velleman is careful to say that the agent doesn't have to justify her actions by consciously apprehending their intelligibility. Intelligibility counts as a higher-order condition of success, while the agent justifiably attends to the specific aims she has while deciding whether or not to perform a certain action. However, as a constitutive aim of all full-blown action, intelligibility is watching over the particular aims and actions the agent decides for. An aim contrary to the higher-level aim of making sense to the agent is an unjustified aim.

Velleman believes only a constitutive aim of action can offer ultimate justification for action and proposes intelligibility as just such an aim. For reasons I cannot explore here, I believe this kind of justification can be offered by the total set of the agent's values, coherently conceived.[104] Assuming this view, there seem to be two ways of conceiving of the justification of a given action or judgment. On the more ambitious version, the agent (or some external assessor) has to evaluate whether her action or judgment coheres with her set of values before acting or judging justifiably. This is the kind of Socratic, overly demanding requirement I am concerned with criticizing all along this chapter. On a more modest interpretation, analogous to Velleman's view, coherence with one's set of values is only a higher-order aim, capable of delegitimizing a particular action or judgment, but that needn't

[103] I say "considered view" because he changed his mind about this. In his 2000b, where the problem is laid out, he takes autonomy to provide the ultimate justification. After criticism from Philip Clark, he changed his view and promoted intelligibility as the ultimate criterion (cf. Velleman, 2000a: 30, fn. 37).

[104] I defend this in chapters 1 and 3. On my view, claims about the ultimate source of justification for action, on the one hand, and claims about the nature of the agent, on the other, converge on a single answer: the agent just is her total set of (coherent) values. It is a modified version of Watson's view in (1982). Velleman briefly considers this kind of coherentist response, in association with Street's work, and seems to find it valid but with some reservations. See his (2009: 126-7, fn. 12).

be consciously entertained by the agent for rational, full-blown action and decision-making.

The problem with this second interpretation is how does the agent find out if her judgment or action coheres with her set of values if she doesn't interrogate it, if she doesn't investigate her values, the sources of justification for her claims and so on. My answer is that she doesn't know and that she shouldn't try to know as well. In the next few paragraphs, I will build a perhaps surprising case for the claim that the agent is often justified in settling on the practical reason her initial reasoning presents her with. In those cases (a large majority, in my view), there is a non-negligible possibility of mismatch between the practical reason she entertains as the conclusion of her practical reasoning, on the one hand, and the practical reason she *would have* were she to follow strictly what is entailed by her total set of values in conjunction with the non-normative facts of the situation, on the other. Since I am claiming that frequently the agent does *not* have a reason to seek total justification via a thorough questioning of her entertained reasons, I am also claiming that she is probably sometimes[105] justified in *not* acting in accordance with what is entailed by her set of values. In short, sometimes she has reason not to be quite herself.[106]

One thing we learn from the literature on non-deliberate acting for reasons is their correct construal of deliberation itself as a kind of action.[107] Being a kind of rational, consciously conducted action, deliberation is subject to criticism and in need for justification. The same is true about deliberation about deliberation, that is, practical reflection about the outcomes of a first instance of practical reflection. As stated in the beginning of this section, we are already limiting our discussion to cases of reflected decision-making. This means the agent is already engaged in the activity of deliberating about what to do. Having reasoned her away through the relevant mental states and external facts, she has settled on a practical conclusion establishing what she has reason to do. She now has the possibility of either doing what she

[105] Why "probably sometimes"? Because whether the mismatch occurs is an empirical matter.

[106] That is, she is justified in doing or judging something that fits less perfectly with her set of values than a possible different action or judgment. See fn. 104 above for a hint at the idea that the valuing agent's identity just is her set of coherent values. Notice as well that the justificatory regress can be stopped at two points. First, as I mentioned above in the text, *if* one actually goes on with subsequent challenges to one's reasons, creating the regress, I claim that it can be stopped once we reach the coherent set of values defining who the agent is. Second, I claim that there is often *no reason* to engage in the regress in the first place. This text is primarily concerned with the second claim. The first claim is defended in chapter 3.

[107] Cf. Arpaly and Schroeder (2012: 210), Railton (2009: 102).

concluded she has reason to do or engaging in second-order deliberation: she may question her conclusion and deliberate about whether to accept it. She may, for instance, ask what justifies the considerations she took to justify her initial conclusion. Is she justified in engaging in this second-order kind of deliberation? The answer is, in most cases, no.

Consider the following story. An agent deliberates and settles on a reason to act. Call the agent A. Call the act of deliberating D. Call the reason she settles on DR (for defeasible reason). And call the action her reason favours performing X.

We are now faced with the question whether what she takes to be a reason (DR) is *sufficient* reason, in the sense that it settles *for the agent* the matter as to what she ought to do. If the agent were to engage in further deliberation (call it D*; for example, look for the entailments of her values and beliefs, look for her dispositions to value and believe, question directly the justification of the reason she had settled on (DR) etc.), it is possible, though not certain, that what she would then conclude to be her considered reason for action (call it CR) would be different from DR.

Engaging in D* is exactly the kind of thing the Socratic model would recommend. But from the way the example was set, she has no reason whatsoever to engage in D*. Being an action just like any other, D* demands justification and the agent can offer none for it. In contrast, there is a perfectly reasonable explanation for why she should perform a competing act, namely action X. For she has reflectively come to the conclusion that she has reason (DR) for action X. Confronted with the possibility of action D*, she may do it or not. Confronted with the possibility of action X, she may do it or not. What differentiates these two courses of action is that the agent has a reason for one (X) and not for the other (D*). What we may conclude from this case is that what stops the potential regress of justification is that the agent has no reason to start it.

5.4. Extending Practical Reflection

If the argument above is sound, an agent is only going to be justified in deliberating about the outcomes of her initial deliberation if presented with a reason to do so. This might be rare, but it does happen. Sometimes we are justified in extending our practical reasoning. I want to explore two kinds of situation in which these

extensions are reasonably called for. I don't claim to have all possible cases covered, but I see them as the two kinds of paradigmatic cases.

The first kind of case I want to consider is that of a challenge from someone else. Suppose you have to decide what to do in a particular situation. You consider what to do reflectively, thinking about the facts of the situation and your mental states at hand (your values, beliefs, desires and intentions that surface in your consciousness as you deliberate). Upon reflection, you settle on a course of action, taking yourself to have reason R to do it. From what we discussed previously, it should follow that you are not justified in questioning further your alleged reason for action, by, for instance, checking its consistency with dispositions to believe and value that you have but didn't consider at first. Suppose now someone notices you taking R to be your reason for action and presents a challenge, maybe demanding more justification from you, maybe outright claiming you are wrong about what you should do. In many such cases of challenge, you will be justified in engaging in further deliberation.

Why is the challenge from others a reason to deliberate further? Someone may doubt that we should give any credit to challenges raised by others. In fact, unless I previously adopt the principle of paying attention to challenges raised by others (maybe only a few others), it is not obvious why I should be bothered. More precisely, why should the course of action I settled for lose in justification when presented with an external challenge, at least if I want to remain within the confines of a broadly internalist picture? It has been suggested that the very notion of a reason is a product for social consumption (Mercier and Sperber 2011), so perhaps the kind of social justification at the base of responding to challenges by others just is what reasons are meant to do. That is possible, but there is another way of explaining the role of challenges in pressing us in the direction of further deliberation, an explanation that fits squarely within the internalist framework.

Challenges are often *de facto* initiators of (further) deliberation. That is, they bring to mind images or ideas related to the topic and relevant to our conclusion regarding what to do. We may, of course, discard them as unwelcome intruders. But, provided we were participating in a conversation in good faith, trying to understand what we were being told, the challenge will have already made its way into our mind. We are already considering it and we may reject it as we may reject just about any other consideration we might be entertaining, regardless of how it managed to pop into our minds. And, if the best reflective decision is reached by striking some kind of

reasoned balance between the beliefs, desires, intentions and values under consideration, the introduction of a new consideration may well destabilize the previous equilibrium and force a new one. In other words, a challenge *de facto* succeeds in broadening the scope of mental states and/or external facts under consideration by the agent and, as such, naturally leads to further deliberation and potentially different conclusions. To disregard the challenge you would need a reason. Absent a reason, you should stick to what you have, which now also includes the challenge. Therefore, challenges do sometimes justifiably lead to greater reflection.

Another kind of situation in which it is reasonable to engage in further deliberation is the case of dilemmas or conflicts. Sometimes reasoning through the mental states and external facts currently available to the agent does not yield a clear answer as to what the agent has most reason to do. It can be that the impasse is unavoidable; there just is no fact of the matter making action A better than action B, as seen from her perspective. But it is natural to suppose that further enquiry, either into the non-normative facts about the case or into one's set of values, beliefs and dispositions, will in many times alter the balance in favour of one or another course of action. It is difficult to specify exactly how much more enquiry is called for. One reasonable policy is (1) to not investigate too much so as to impart one's ability to implement one's (other) practical interests; (2) and stop investigating once the balance has favoured one of the competing courses of action. At this point, the agent will not have reason to investigate further, for the same reasons as the one's discussed in the previous section.[108]

A related set of cases is that of big, transformative decisions, such as changing one's career, getting married or having a child. These are cases in which it seems fitting to step back and reflect on the what kind of person one is or wants to become.[109] However, it is misleading to count them as cases of further deliberation in the sense I am considering here. These are certainly important decisions, which demand a great deal of reflection, but they don't require the kind of *further* reflection advocated by the Socratic model. It is one thing to consider what kind of person I am or want to be and how going for such transformative experiences helps or doesn't

[108] At the moment when it becomes clear that further investigation is not helping any of the competing courses of action and is risking damaging one's capacity to do other things, the situation turns into that of an arbitrary decision. Both courses of action being permitted, there is no clear reason to favour one over the other.

[109] I cannot do justice to the complexity of this theme here. For a recent and influential discussion, see (Paul 2014).

help implement my ambitions. It is another thing to engage in critical reflection about the foundations of my commitments. The Socratic injunction is to fully examine the purported reasons I have to believe and act, because it is expected that only reasons endorsed after going through this critical scrutiny are really reasons for action and belief. Thinking about having or not a child doesn't ask me to challenge the justification for my values. It asks me to entertain a scenario and see if my values move in its direction or not.

5.5. Problems with full-analysis

What we have discussed so far shows that all-out reflection is not necessary for leading a life responsive to one's practical reasons. Deliberation about one's practical reasons is an action just like any other, an action one may or may not be justified in engaging in, depending on certain features of the context of action. The reasons against full-reflection I have given so far have been reasons to stick to a condition falling short of full reflective self-analysis. What I wish to do in this final section is discuss what might happen if one *does* go for full reflection. What do we get from engaging in such a pursuit?

An initial worry related to the project of full self-examination has to do with its feasibility. Given that we are human beings, with limited mental storage capacities, we shouldn't go about acquiring every piece of knowledge about ourselves we are capable of and we shouldn't go about drawing all true inferences from our beliefs and intentions we are capable of. The consequence of this would be unjustified mind cluttering, which would inevitably stand in the way of our practical interests.[110]

A second worry relates to the coherentist model of justification underlying metaethical constructivism. In coherentism, a judgment is justified roughly to the extent that it fits into a web of mutually justifying beliefs. For any given (normative) judgment, we may assess its justification according to a (coherent) set of beliefs. Although our previous considerations recommended against it, nothing prevents the assessor from moving along a challenging regress. But coherentism has no problem with this. The coherentist will be able to stop the regress at some given higher level,

[110] Cf. Harman (1986: 12), where he presents his Clutter Avoidance principle. Harman groups the avoidance of mind-cluttering inferences together with the idea that we need a reason to be justified in challenging a current belief (which I discussed in the previous two sections) under the label: Principle of Conservatism (1986: 116).

at the moment in which the belonging of the assessed normative judgment to the web of beliefs of reference is proven or disproven.

Above, we saw there is often no reason to move along this justificatory chain. But now, assuming we engage in it, coherentism has a principled and traditional answer to stopping the regress. However, the answer only works if we are assessing one judgment or belief at a time (or a subset of judgments or beliefs). But the kind of examination associated with the philosophical way of life easily turns into a general challenge to one's value system. Placing all one's values or beliefs under suspicion with the expectation of establishing a firmer set afterwards has enormously disruptive potential. This is so because, while it is possible to offer another judgment as a justification for a first judgment, there is nothing, from the point of view of a coherentism about values, that one can offer as a justification for one's whole set of values. If challenged in tandem, one's values – in fact one's identity – reveal their arbitrariness. Of course, if one adopts a detached or theoretical stance in relation to one's mental states, there might be no problem with revealing the arbitrary, genealogical processes that caused one to possess them. But from the engaged, first-personal perspective of the agent[111] looking for a justification for her normative judgments, the sense that they are arbitrary and the lack of justification to offer upon their being challenged as a whole may well have negative consequences for one's capacity to move on with one's practical, valuing life.[112][113]

---

[111] In referring to theoretical versus practically engaged stances or attitudes, I am adopting Moran's terminology (2001). From the theoretical stance, nothing "really matters". But that is not what matters, or better, that is not *how* things matter (they matter from the engaged point of view). I borrow the talk about what "really matters" from Street (Street 2017).

[112] This difficulty with moving on with one's life is probably part of what Nietzsche had in mind with his motto that "Truth is terrible". Cf. Leiter (2018) for discussion. Williams' reflections on the difficulties and problems associated with uncovering the contingent history of our outlooks are also relevant here. Cf. (Williams 2006).

[113] Valerie Tiberius' theory faces a similar problem: "[There] is a kind of philosophical reflection that seeks an ultimate foundation for our value commitments and the reasons we take ourselves to have for them, where we do not take for granted the value of anything. The Reflective Wisdom Account might seem vulnerable to this kind of reflection, because it does not supply an inherently normative foundation for the authority or legitimate force of the reasons we have to change our habits of thought or act on our values" (2008: 182). Setting aside this general, philosophical reflection, her view is quite close to mine: "The meta-ethics suggested by the Reflective Wisdom Account locates the source of normativity for its prescriptive claims in the stable network of commitments of a reflective agent who has a concern to live a life she can endorse. When we engage in reflection on our commitments, some of them must be taken for granted in order to reflect critically on others (though, of course, not necessarily the same ones all the time). In particular, the underlying commitment to living a life that meets our reflective standards is always taken for granted and cannot itself be justified. On this picture, then, our reasons for pursuing the things we value and for living up to our standards ultimately derive from our concern to live well and from other particular commitments we have" (*Idem*: 183). The main difference between our views is that she places the value of living a life that meets one's reflective

Someone might object that knowledge has intrinsic value and that (self-) knowledge is worth pursuing for its own sake. On this view, acquiring information about one's own values and their entailments is a worthwhile enterprise, which may not be the best thing to do at all times, but counts certainly as justified conduct from time to time. In reply to this line of reasoning, we must first notice that it is not the value of knowledge *per se* that is at stake here, but rather whether we should or shouldn't engage in the pursuit of self-knowledge or self-examination. As an activity, it ought to be assessed in contrast to other activities. What the previous sections have shown is that this pursuit is often detrimental to the agent's overall practical interests, by unjustifiably deviating her from what she has most reason to do. This conclusion could have been different if we presupposed some kind of good life, conceived in realist terms, whose fulfilment was inextricably connected to self-examination. This is probably true of Socrates' philosophy, but goes against the metaethical constructivism I have been assuming all along. Furthermore, as noted in the previous paragraph, within the coherentist approach characteristic of constructivism, a full self-examination has the potential to shake the grounds of the agent's values, without having anything to offer in substitution. This means that, depending on how it is undertaken – if not piecemeal – a project of full self-analysis is likely to lead straight into nihilism or suffering.[114]

Absence of a realist conception of the good life attached to self-knowledge, frequent conflict with one's practical interests and potential disruption of one's system of values are the three main reasons why living a life in the pursuit for Socratic self-knowledge is in all likelihood a bad idea. Coupled with the lack of reasons to progress, from actual cases of deliberation, into that situation of Socratic reflection, these reasons make the philosophical way of life a good candidate for misery.

---

standards above all other values and assesses the other values according to that higher value. I see no reason to posit such hierarchy.

[114] A point I don't explore in the text is the possibility that seeking self-knowledge actually alters the content of one's mental states, so that the "self" to be known becomes different upon reflection. This point is acutely explored by Moran (2001: 59 and chapter 5). Tiberius accepts this point from Moran (2008: 116).

Conclusion

Having discussed so many different things, it is possible that I left the reader without a clear grasp on the argumentative thread of the dissertation. I will attempt to remedy this situation in this Conclusion. The reader will have noticed that the first three chapters are primarily dedicated to the theoretical presentation and grounding of my preferred version of metaethical constructivism. Conversely, the two last chapters explore implications and extensions of the view.

The view in question is of course Humean constructivism with a few caveats. My main caveats to Street's Humean constructivism were discussed in chapter 1. There, I took issue with the formalistic presentation of her view and with the way constructivism in general typically construes the role of the procedure in the justification of practical reasons. On my view, a defensible version of metaethical constructivism has to accommodate the role played by the contingent values held by particular agents in justification making it, thus, impossible to understand constructivist in purely formal terms. Instead, I argued that it is our specific values that have certain rational and logical features and that we shouldn't be looking beyond them to find the normative materials for practical justification.

In light of this, I also reconceived the role of the constructive procedure. For me, reflective scrutiny is more of an epistemological than of a constructive nature. Its task is to discover what are the values the agent already holds and what are their relevant entailments. I stress "relevant" entailments, because, as we saw in the last chapter, it wouldn't make sense to require of agents that they draw all possible consequences from the values and beliefs they hold. Perhaps surprisingly, what I wish to have provided convincing arguments for is the claim that in fact agents should resist the temptation to overly reflect a lot sooner than traditional arguments pointing to the risk of mind cluttering have suggested. If correct, my argumentation allegedly shows that the unexamined life is in many cases the best life.

This conclusion apparently contradicts the view, defended especially in chapter 3, connecting the agent's set of values to her identity and her identity to the justification of practical reasons. Recall that in chapter 3 I argued that an important part of the justification an agent has in taking a certain consideration to be a reason for her lies in the fact that this consideration is based on a value that is constitutive of the agent she is. But then in chapter 5 I claim that it is often the case the we should

not try to find out who we are and, hence, what our ultimate values are. This last claim, coupled with the point that reasons stem from deeply held values, seems contradictory, in that it *prima facie* suggests we should *not* do what we have most reason to do.

The contradiction is illusory, however, and dissolving its impression requires bearing in mind that authenticity (acting as one truly is) is not the only source of practical justification. The argument is chapter 5 is meant to have shown under what circumstances other kinds of considerations defeat the search for a more authentic normative judgment. In those cases, what we have most reason to do is to not be quite ourselves.

An important limitation of this dissertation is that it doesn't engage very much with competing theories of practical normativity. There is little or no discussion of non-naturalist realism, naturalist realism, error theory and non-cognitivism, just to mention some of the main competitors in the complex and intricate metaethical literature. In fact, I was more concerned with showing that constructivism should be seen as a plausible, defensible view than with showing that it is necessarily the best theory in metaethics. At some interesting moments in the text, however, a few points of connection with other traditions appeared in often-unexpected ways. That is the case of my interpretation of the constructivist procedure as epistemological, very much in the realist spirit. It is probably also the case of some of my remarks about moral disagreement beyond the communication of reasons: many readers will find parallels between my arguments there and traditional arguments in non-cognitivism (e.g. the idea of disagreement in attitude, to be found in Stevenson).

If seen from the internal perspective of metaethics, the dissertation has had a rather limited scope, as it didn't discuss other traditions, seen from the broader philosophical perspective, the scope was perhaps too large. Part of the explanation for this lies in the very nature of metaethics: at the same time a subpart of ethics and a field gathering considerations from different fields of philosophy. So I have had to discuss topics in epistemology, philosophy of language, philosophy of mind and action, experimental philosophy and even science in ways that maybe other subfields of philosophy would not have to. I consider this feature of metaethics one of its most appealing ones. It is a big challenge to claim just about anything in metaethics, given the consequences it has for other aspects of (moral) reality. But the parts of reality are

themselves deeply interconnected, so I hope metaethics can be of some help in our attempt at better understanding them.

References

Arpaly, N., and T. Schroeder. 2012. "Deliberation and Acting for Reasons." *Philosophical Review* 121 (2): 209–39. https://doi.org/10.1215/00318108-1539089.

Bagnoli, Carla. 2017. "Constructivism in Metaethics." In *The Stanford Encyclopedia of Philosophy*, Edward N. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/constructivism-metaethics/.

Bayne, Tim, and Elisabeth Pacherie. 2007. "Narrators and Comparators: The Architecture of Agentive Self-Awareness." *Synthese* 159 (3): 475–91. https://doi.org/10.1007/s11229-007-9239-9.

Beebe, James R. 2014. *Advances in Experimental Moral Psychology*. Edited by Hagop Sarkissian and Jennifer Cole Wright. *Advances in Experimental Moral Psychology*. London, UK: Bloomsbury Academic. https://doi.org/10.5040/9781472594150.

Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality : A Treatise in the Sociology of Knowledge*. New York: Doubleday & Company.

Boghossian, Paul. 2011. "Three Kinds of Relativism." In *A Companion to Relativism*, edited by Steven D. Hales, 53–69. Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781444392494.ch3.

Bratman, Michael E. 1999. "Identification, Decision, and Treating as a Reason." In *Faces of Intention: Selected Essays on Intention and Agency (Cambridge Studies in Philosophy)*, 185–206. Cambridge: Cambridge University Press.

———. 2007. "Introduction." In *Structures of Agency : Essays*, 1–18. Oxford: Oxford University Press.

Brogaard, Berit. 2008. "Moral Contextualism and Moral Relativism." *Philosophical Quarterly* 58 (232): 385–409. https://doi.org/10.1111/j.1467-9213.2007.543.x.

Broome, John. 1999. "Normative Requirements." *Ratio* 12 (4): 398–419. https://doi.org/10.1111/1467-9329.00101.

———. 2013. *Rationality Through Reasoning*. *Rationality Through Reasoning*. West Sussex, UK: Wiley Blackwell. https://doi.org/10.1002/9781118609088.

Bukoski, Michael. 2017. "Self-Validation and Internalism in Velleman's Constitutivism." *Philosophical Studies* 174 (11): 2667–86. https://doi.org/10.1007/s11098-016-0804-2.

———. 2018. “Korsgaard’s Arguments for the Value of Humanity.” *Philosophical Review* 127 (2): 197–224. https://doi.org/10.1215/00318108-4326604.

Cappelen, Herman, and John Hawthorne. 2009. *Relativism and Monadic Truth*. *Relativism and Monadic Truth*. Oxford: Oxford University Press UK. https://doi.org/10.1093/acprof:oso/9780199560554.001.0001.

Chalmers, David J. 2011. “Verbal Disputes.” *Philosophical Review* 120 (4): 515–66. https://doi.org/10.1215/00318108-1334478.

Chang, Ruth. 2009. “Voluntarist Reasons and the Sources of Normativity.” In *Reasons for Action*, 9780521877:243–71. Cambridge University Press. https://doi.org/10.1017/CBO9780511720185.012.

Cooper, John M. 2012. *Pursuits of Wisdom: Six Ways of Life in Ancient Philosophy from Socrates to Plotinus*. New Jersey: Princeton University Press.

Darwall, Stephen, Allan Gibbard, and Peter Railton. 1992. “Toward Fin de Siecle Ethics: Some Trends.” *The Philosophical Review* 101 (1): 115. https://doi.org/10.2307/2185045.

Dorsey, Dale. 2018. “A Perfectionist Humean Constructivism.” *Ethics* 128 (3): 574–602. https://doi.org/10.1086/696152.

Driver, Julia. 2017. “Contingency and Constructivism.” In *Reading Parfit*, 172–88. Routledge. https://doi.org/10.4324/9781315225531-10.

Enoch, David. 2006. “Agency, Shmagency: Why Normativity Won’t Come from What Is Constitutive of Action.” *The Philosophical Review* 115 (2): 169–98. https://doi.org/10.1215/00318108-2005-014.

———. 2009. “Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?” *Philosophical Explorations*. https://doi.org/10.1080/13869790903067683.

———. 2011. “Shmagency Revisited.” In *New Waves in Metaethics*, 208–33. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230294899_11.

Ferrero, Luca. 2009. “Constitutivism and the Inescapability of Agency.” *Oxford Studies in Metaethics* 4: 303–33. https://doi.org/10.1093/acprof:oso/9780199653492.001.0001.

Frankfurt, Harry G. 1988. “Freedom of the Will and the Concept of a Person.” In *The Importance of What We Care About*, 11–25. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511818172.003.

Fricker, Miranda. 2010. “The Relativism of Blame and Williams’s Relativism of

Distance." *Proceedings of the Aristotelean Society, Supplementary Volumes*. https://doi.org/10.1111/j.1467-8349.2010.00190.x.

———. 2016. "What's the Point of Blame? A Paradigm Based Explanation." *Nous* 50 (1): 165–83. https://doi.org/10.1111/nous.12067.

Gaut, Berys. 1997. "The Structure of Practical Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 161–88. New York: Oxford University Press.

Gibbard, Allan. 1999. "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics* 110 (1): 140–64. https://doi.org/10.1086/233207.

Hadot, Pierre. 1995. *Philosophy as a Way of Life: Spiritual Exercises from Socrates to Foucault*. Edited by Arnold Davidson. Translated. Blackwell Publishers.

Harman, Gilbert. 1986. *Change in View : Principles of Reasoning*. Cambridge, MA: MIT Press.

Hume, David. 2011. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. *Vol. 1: Texts*. A Clarendon Press Publication. https://doi.org/10.1093/actrade/9780199596331.book.1.

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

Khoo, Justin, and Joshua Knobe. 2016. "Moral Disagreement and Moral Semantics." *Nous* 52 (1): 109–43. https://doi.org/10.1111/nous.12151.

Kölbel, Max. 2004. "Faultless Disagreement." *Proceedings of the Aristotelean Society* 104 (1): 53–73. https://doi.org/10.1111/j.0066-7373.2004.00081.x.

Kolodny, Niko. 2005. "Why Be Rational?" *Mind* 114 (455): 509–63. https://doi.org/10.1093/mind/fzi509.

Kornblith, Hilary. 2010. "What Reflective Endorsement Cannot Do." *Philosophy and Phenomenological Research* 80 (1): 1–19. https://doi.org/10.1111/j.1933-1592.2009.00307.x.

Korsgaard, Christine M. 1989. "Morality as Freedom." In *Kant´s Practical Philosophy Reconsidered*, edited by Yovel Yirmiyahu, 23–48. Boston: Kluwer.

———. 1996a. "Kant's Formula of Humanity." In *Creating the Kingdom of Ends*, 106–32. New York: Cambridge University Press.

———. 1996b. *The Sources of Normativity*. *The Sources of Normativity*. Cambridge University Press. https://doi.org/10.1017/cbo9780511554476.

———. 1997. “The Normativity of Instrumental Reason.” In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut. Clarendon press.

———. 2003. “Christine M. Korsgaard: Internalism and the Sources of Normativity (Interview with C. Korsgaard).” In *Constructions of Practical Reason*, edited by Herlinde Pauer-Studer, 50–69. Stanford University Press.

———. 2009. *Self-Constitution: Agency, Identity, and Integrity*. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552795.001.0001.

Kraut, Richard. 2007. “The Examined Life.” In *A Companion to Socrates*, edited by Sara Ahbel-Rappe and R. Kamtekar, 228–42. Oxford: Blackwell Publishers.

Langton, Rae. 2007. “Objective and Unconditioned Value.” *Philosophical Review* 116 (2): 157–85. https://doi.org/10.1215/00318108-2006-034.

Leiter, Brian. 2018. “The Truth Is Terrible.” *Journal of Nietzsche Studies* 49 (2): 151–73.

Levy, Yair. 2019. “Constructivism and the Problem of Normative Indeterminacy.” *Journal of Value Inquiry* 53 (2): 243–53. https://doi.org/10.1007/s10790-018-9664-7.

List, Christian. 2016. “What Is It Like to Be a Group Agent?” *Nous* 52 (2): 295–319. https://doi.org/10.1111/nous.12162.

List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199591565.001.0001.

MacFarlane, John. 2005. “Making Sense of Relative Truth.” *Proceedings of the Aristotelean Society* 105 (1): 305–23. https://doi.org/10.1111/j.0066-7373.2004.00116.x.

———. 2007. “Relativism and Disagreement.” *Philosophical Studies* 132 (1): 17–31. https://doi.org/10.1007/s11098-006-9049-9.

———. 2009. “Nonindexical Contextualism.” *Synthese* 166 (2): 231–50. https://doi.org/10.1007/s11229-007-9286-2.

———. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. New York: Oxford University Press.

Mason, Elinor. 2019. *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780198833604.001.0001.

McHugh, Conor, and Jonathan Way. 2018a. "What Is Good Reasoning?" *Philosophy and Phenomenological Research* 96 (1): 153–74. https://doi.org/10.1111/phpr.12299.

———. 2018b. "What Is Reasoning?" *Mind* 127 (505): 167–96. https://doi.org/10.1093/mind/fzw068.

Mercier, Hugo, and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory." *The Behavioral and Brain Sciences* 34 (2): 57–74; discussion 74-111. https://doi.org/10.1017/S0140525X10000968.

Moran, Richard. 2001. *An Essay on Self-Knowledge*. Princenton: Princeton University Press. http://www.jstor.org/stable/j.ctt7sgs5.

Nelkin, Dana K. 2000. "Two Standpoints and the Belief in Freedom." *The Journal of Philosophy* 97 (10): 564. https://doi.org/10.2307/2678468.

Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198717959.001.0001.

Plato. 1997. *Plato: Complete Works*. Edited by John Cooper and D.S. Hutchison. Cooper. Indianapolis: Hackett. https://doi.org/10.5962/bhl.title.17593.

Plunkett, David, and Timothy Sundell. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers Imprint* 13.

Railton, Peter. 2009. "Practical Competence and Fluent Agency." In *Reasons for Action*, edited by David Sobel and Steven Wall, 81–115. New York: Cambridge University Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass: Belknap Press of Harvard University Press.

———. 1980. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77 (9): 515. https://doi.org/10.2307/2025790.

Richard, Mark. 2008. *When Truth Gives Out*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199239955.001.0001.

Rosati, Connie S. 2016. "Agents and 'Shmagents.'" In *Oxford Studies in Metaethics:11*, 182–213. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198784647.003.0008.

Sarkissian, Hagop, John Park, David Tien, Jennifer Cole Wright, and Joshua Knobe. 2011. "Folk Moral Relativism." *Mind and Language* 26 (4): 482–505. https://doi.org/10.1111/j.1468-0017.2011.01428.x.

Sayre-McCord, Geoffrey. 1996. "Coherentist Epistemology and Moral Theory." In

Moral Knowledge? New Readings in Moral Epistemology, edited by Walter Sinnott-Armstrong and Mark Timmons, 137–89. New York: Oxford University Press.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap/Harvard University Press.

Schafer, Karl. 2014. "Constructivism and Three Forms of Perspective-Dependence in Metaethics." *Philosophy and Phenomenological Research* 89 (1): 68–101. https://doi.org/10.1111/j.1933-1592.2012.00633.x.

Scheffler, Samuel. 2011. "Valuing." In *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199753673.003.0002.

Shafer-Landau, Russ. 2003. *Moral Realism*. New York: Oxford University Press. https://doi.org/10.1093/0199259755.001.0001.

Silverstein, Matthew. 2015. "The Shmagency Question." *Philosophical Studies* 172 (5): 1127–42. https://doi.org/10.1007/s11098-014-0340-x.

Snowdon, Paul F. 2018. "Review of Galen Strawson, The Subject of Experience, Oxford University Press, 2017." Available at Https://Ndpr.Nd.Edu/News/the-Subject-of-Experience/. 2018.

Southwood, Nicholas. 2018. *Constructivism About Reasons*. Edited by Daniel Star. *The Oxford Handbook of Reasons and Normativity*. Vol. 1. New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199657889.013.16.

Stojanovic, Isidora. 2017. "Metaethical Relativism." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett. New York: Routledge. https://doi.org/10.4324/9781315213217.

Strawson, Galen. 2004. "Against Narrativity." *Ratio* 17 (4): 428–52. https://doi.org/10.1111/j.1467-9329.2004.00264.x.

———. 2017. *The Subject of Experience*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198777885.001.0001.

Strawson, P.F. 1974. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. New York: Routledge.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 127 (1): 109–66. https://doi.org/10.2307/4321684.

———. 2008a. "Constructivism About Reasons." *Oxford Studies in Metaethics:*

Volume 3 3: 207–45. https://doi.org/10.1093/acprof:oso/9780199606375.001.0001.

———. 2008b. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About." *Philosophical Issues* 18 (1): 207–28. https://doi.org/10.1111/j.1533-6077.2008.00145.x.

———. 2009. "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters." *Philosophical Issues* 19 (1): 273–98. https://doi.org/10.1111/j.1533-6077.2009.00170.x.

———. 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84. https://doi.org/10.1111/j.1747-9991.2009.00280.x.

———. 2012. "Coming to Terms with Contingency: Humean Constructivism about Practical Reason." In *Constructivism in Practical Philosophy*, edited by Jimmy Lenman and Yonatan Shemmer. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199609833.003.0003.

———. 2017. "Nothing 'Really' Matters, but That's Not What Matters." In *Does Anything Really Matter?: Essays on Parfit on Objectivity*, edited by Peter Singer, 121–48. Oxford: Oxford University Press.

Theunissen, L. Nandi. 2018. "Must We Be Just Plain Good? On Regress Arguments for the Value of Humanity." *Ethics* 128 (2): 346–72. https://doi.org/10.1086/694273.

Tiffany, Evan. 2012. "Why Be an Agent?" *Australasian Journal of Philosophy* 90 (2): 223–33. https://doi.org/10.1080/00048402.2011.605792.

Velleman, J. David. 2000a. "Introduction." In *The Possibility of Practical Reason*, 1–31. Oxford: Oxford University Press.

———. 2000b. "The Possibility of Practical Reason." In *The Possibility of Practical Reason*, 170–99. Oxford: Oxford University Press.

———. 2000c. "What Happens When Someone Acts." In *The Possibility of Practical Reason*, 123–43. Oxford: Oxford University Press.

———. 2009. *How We Get Along*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808296.

———. 2013. *Foundations for Moral Relativism*. *Foundations for Moral Relativism*. Open Book Publishers. https://doi.org/10.11647/OBP.0029.

Vlastos, Gregory. 1994. *Socratic Studies*. Edited by Myles Burnyeat. New York: Cambridge University Press.

Watson, Gary. 1982. “Free Agency.” In *Free Will*, 95–110. Oxford: Oxford University Press.

Way, Jonathan. 2010. “The Normativity of Rationality: The Normativity of Rationality.” *Philosophy Compass* 5 (12): 1057–68. https://doi.org/10.1111/j.1747-9991.2010.00357.x.

Williams, Bernard. 1979. “Internal and External Reasons.” In *Rational Action*, edited by Ross Harrison, 101–13. Cambridge: Cambridge University Press.

———. 1995. “Internal Reasons and the Obscurity of Blame.” In *Making Sense of Humanity*, 35–45. Cambridge, MA: Cambridge University Press.

———. 2006. “Philosophy as a Humanistic Discipline.” In *Philosophy as a Humanistic Discipline*, 180–99. New Jersey: Princeton University Press.

Wright, Crispin. 2008. “Relativism about Truth Itself: Haphazard Thoughts about the Very Idea.” In *Relative Truth*, 157–86. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199234950.003.0007.

Zeman, Dan. 2017. “Contextualist Answers to the Challenge from Disagreement.” *Phenomenology and Mind* 0 (12): 62–73. https://doi.org/10.13128/Phe_Mi-21106.